

SEARCH REQUEST FORM

Requestor's Name: _____ Serial Number: _____
Date: _____ Phone: _____ Art Unit: _____

Search Topic:

Please write a detailed statement of search topic. Describe specifically as possible the subject matter to be searched. Define any terms that may have a special meaning. Give examples or relevant citations, authors, keywords, etc., if known. For sequences, please attach a copy of the sequence. You may include a copy of the broadest and/or most relevant claim(s).

BEST AVAILABLE COPY

STAFF USE ONLY

Date completed: 12-09-03
Searcher: Beverly C4994
Terminal time: 20
Elapsed time: _____
CPU time: _____
Total time: 25
Number of Searches: _____
Number of Databases: 1

Search Site
____ STIC
____ CM-1
____ Pre-S
Type of Search
____ N.A. Sequence
____ A.A. Sequence
____ Structure
____ Bibliographic

Vendors
____ IG
____ STN
____ Dialog
____ APS
____ Geninfo
____ SDC
____ DARC/Questel
____ Other CGN

THIS PAGE BLANK (USPTO)



STIC Search Report

Biotech-Chem Library

STIC Database Tracking Number: 109333

To: Manjunath N Rao
Location: cm1/10a11 & 10d01
Art Unit: 1652
Wednesday, December 10, 2003

Case Serial Number: 09/980729

From: Beverly Shears
Location: Biotech-Chem Library
CM1-1E05
Phone: 308-4994

beverly.shears@uspto.gov

Search Notes

THIS PAGE BLANK (USPTO)

109333

From: Rao, Manjunath N.
Sent: Monday, December 01, 2003 9:36 AM
To: STIC-Biotech/ChemLib
Subject: Sequence search request for 09/980729

From: Manjunath N. Rao
Art Unit 1652, Room 10A11
Mail Box in Room 10D 01
Phone: 306-5681

Date: 12-1-03

Please search the following as soon as possible for application with serial number

09/980,729

1. SEQ ID NO: 10 against all commercial nucleic acid databases, issued patents/published applications database and pending application database. Please provide a print of all results
2. SEQ ID NO: 5 against all commercial amino acid databases, issued patents/published applications database and pending application database. Please provide a print of all results.

If you have any questions please call me at the above phone number.

Thanks

Manjunath N. Rao, Ph.D.
Biotechnology Patent Examiner
Art Unit 1652, Room 10A11
Mail Box in 10D01
Crystal Mall 1, USPTO.

Searcher: _____
Phone: _____
Location: _____
Date Picked Up: _____
Date Completed: _____
Searcher Prep/Review: _____
Clerical: _____
Online time: _____

TYPE OF SEARCH:
NA Sequences: _____
AA Sequences: _____
Structures: _____
Bibliographic: _____
Litigation: _____
Full text: _____
Patent Family: _____
Other: _____

VENDOR/COST (where applic.)
STN: _____
DIALOG: _____
Questel/Orbit: _____
DRLink: _____
Lexis/Nexis: _____
Sequence Sys.: _____
WWW/Internet: _____
Other (specify): _____

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 5, 2003, 09:42:55 ; Search time 18 Seconds

(without alignments)
1382.063 Million cell updates/sec

Title: US-09-980-729B-5

Perfect score: 2798

Sequence: 1 MSKMTSALLILQLSCYFSS.....KCFLLSCQKFKTRKIEKRE 529

Scoring table: BIOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_41.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|--------|-------------|--------|----|-------------|
| 1 | 2602.5 | 93.0 | 528 | 1 | UDB4_HUMAN |
| 2 | 2369.5 | 84.7 | 528 | 1 | UDB7_MACFA |
| 3 | 2330 | 83.3 | 529 | 1 | UDB7_HUMAN |
| 4 | 2297 | 82.1 | 529 | 1 | UDB9_MACFA |
| 5 | 2292 | 81.9 | 529 | 1 | UDB8_HUMAN |
| 6 | 2281.5 | 81.5 | 528 | 1 | UDBA_HUMAN |
| 7 | 2253 | 80.5 | 529 | 1 | UDBS_HUMAN |
| 8 | 2169.5 | 77.5 | 530 | 1 | UDBF_HUMAN |
| 9 | 2158.5 | 77.1 | 530 | 1 | UDBF_HUMAN |
| 10 | 2100.5 | 75.1 | 530 | 1 | UDBF_MACFA |
| 11 | 1991.5 | 71.2 | 530 | 1 | UDBE_RABIT |
| 12 | 1973.5 | 70.5 | 529 | 1 | UDBI_RABIT |
| 13 | 1938.5 | 69.3 | 523 | 1 | UDBG_RABIT |
| 14 | 1938 | 69.3 | 531 | 1 | UDBD_RABIT |
| 15 | 1930.5 | 69.0 | 530 | 1 | UDBC_RAT |
| 16 | 1904.5 | 68.1 | 530 | 1 | UDB5_MOUSE |
| 17 | 1894.5 | 67.7 | 530 | 1 | UDB3_RAT |
| 18 | 1856.5 | 66.4 | 530 | 1 | UDB2_RAT |
| 19 | 1840.5 | 65.8 | 530 | 1 | UDB6_RAT |
| 20 | 1777.5 | 63.5 | 530 | 1 | UDB8_RAT |
| 21 | 1761.5 | 63.0 | 527 | 1 | UDAI_RAT |
| 22 | 1593 | 56.9 | 502 | 1 | UDCI_RABIT |
| 23 | 1201.5 | 42.9 | 531 | 1 | UDIS_RAT |
| 24 | 1196.5 | 42.8 | 533 | 1 | UD12_RAT |
| 25 | 1188.5 | 42.5 | 533 | 1 | UD12_RAT |
| 26 | 1158 | 41.4 | 535 | 1 | UD12_MOUSE |
| 27 | 1137.5 | 40.7 | 531 | 1 | UD13_RAT |
| 28 | 1137.5 | 40.7 | 533 | 1 | UD11_HUMAN |
| 29 | 1135.5 | 40.6 | 535 | 1 | UD15_RAT |
| 30 | 1128 | 40.3 | 534 | 1 | UD15_HUMAN |
| 31 | 1126 | 40.2 | 534 | 1 | UD14_HUMAN |
| 32 | 1124 | 40.2 | 534 | 1 | UD13_HUMAN |
| 33 | 1120.5 | 40.0 | 530 | 1 | UD18_HUMAN |

| | | | | | | |
|----|--------|------|-----|---|------------|--------------------|
| 34 | 1120.5 | 40.0 | 530 | 1 | UD19_HUMAN | w06566 homo sapien |
| 35 | 1109.5 | 39.7 | 532 | 1 | UD14_RABIT | Q28612 oryctolagus |
| 36 | 1099.5 | 39.3 | 530 | 1 | UD1A_HUMAN | Q9haw8 homo sapien |
| 37 | 1098.5 | 39.3 | 530 | 1 | UD12_HUMAN | P36509 homo sapien |
| 38 | 1094.5 | 39.1 | 530 | 1 | UD17_HUMAN | Q9haw7 homo sapien |
| 39 | 1094 | 39.1 | 520 | 1 | UD17_MOUSE | Q62452 mus musculu |
| 40 | 1081.5 | 38.7 | 532 | 1 | UD16_HUMAN | P19274 homo sapien |
| 41 | 1078.5 | 38.5 | 531 | 1 | UD17_RAT | Q64633 ratu |
| 42 | 1077.5 | 38.5 | 530 | 1 | UD18_RAT | Q64634 ratu |
| 43 | 1066 | 38.1 | 531 | 1 | UD16_MOUSE | Q64435 mus musculu |
| 44 | 1045.5 | 37.4 | 529 | 1 | UD16_RAT | P08430 ratu |
| 45 | 1045 | 37.3 | 531 | 1 | UD16_RABIT | Q28611 oryctolagus |

ALIGNMENTS

RESULT 1
ID: UDB4_HUMAN STANDARD; PRT; 528 AA.
AC P06133; 060731; 060867; 075614; P36538;
DT 01-JAN-1988 (Rel. 06, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE UDP-glucuronosyltransferase 2B4 precursor, microsomal (EC 2.4.1.17)
DE (UDPGLT) (Hydroxycholelic acid) (HUG25) (UDPGLT-1).
GN UGT2B4 OR UGT2B11.
OS Homo sapiens (human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RX MEDLINE=87241362; PubMed=3109396;
RA Jackson M.R., McCarthy L.R., Harding D., Wilson S., Coughtrie M.W.H.,
RA Burchell B.;
RT "Cloning of a human liver microsomal UDP-glucuronosyltransferase
RT cDNA.";
RL Biochem. J. 242:581-588(1987).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RX MEDLINE=93326164; PubMed=8333863;
RA Jin C.-Y., Miners J.O., Lillywhite K.J., McKenzie P.I.;
RT "cDNA cloning and expression of two new members of the human liver
RT UDP-glucuronosyltransferase 2B subfamily.";
RL Biochem. Biophys. Res. Commun. 194:496-503(1993).
RN [3]
RP SEQUENCE FROM N.A., VARIANT GLU-458, AND CHARACTERIZATION.
RX MEDLINE=99303261; PubMed=10376768;
RA Levesque E., Beaulieu M., Hum D.W., Belanger A.;
RT "Characterization and substrate specificity of UGT2B4 (E458): a
RT UDP-glucuronosyltransferase encoded by a polymorphic gene.";
RL Pharmacogenetics 9:207-216(1999).
RN [4]
RP SEQUENCE FROM N.A., AND VARIANTS LEU-109 AND LEU-396.
RA McKenzie P.I.;
RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.
RN [5]
RP SEQUENCE FROM N.A.
RA Riedy M., Miller A.;
RT "Genomic organization and structure of the UGT2B gene complex at human
RT chromosome 4q31.";
RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
RN [6]
RP FUNCTION: UDPGLTs are of major importance in the conjugation and
RP subsequent elimination of potentially toxic xenobiotics and
RP endogenous compounds. This isozyme is active on polihydroxylated
RP estrogens (such as estradiol, 4-hydroxyestrone and 2-hydroxyestradiol)
RP and xenobiotics (such as 4-methylumbelliferone, 1-naphthol, 4-
RP nitrophenol, 2-aminophenol, 4-hydroxybiphenyl and menthol). It is
RP capable of 6 alpha-hydroxyglucuronidation of hydroxycholelic acid.
CC -!- CATALYTIC ACTIVITY: UDP-glucuronate + acceptor = UDP + acceptor

THIS PAGE BLANK (USPTO)

| | |
|---------|--|
| DR | EMBL; Y00317; CA68415.1; -; |
| DR | EMBL; AF064200; AAC95002.1; -; |
| DR | EMBL; AJ005162; CAA0636.1; -; |
| DR | EMBL; AF081793; AAC3272.1; -; |
| DR | EMBL; AF135416; AAF70145.1; -; |
| DR | PIR; JN0619; JN0619. |
| DR | GeneW; HGNC:12553; UGT2B4. |
| DR | MIM; 600067; -; |
| DR | GO; GO:0005792; C:microsome; NAS. |
| DR | GO; GO:0006711; P:steroid catabolism; IDA. |
| DR | GO; GO:0006805; P:xenobiotic metabolism; IDA. |
| DR | InterPro; IPR002213; UDP_gluco_trans. |
| DR | Pfam; PF00201; UDPE1; 1. |
| DR | PROSITE; PS00375; UDPE1; 1. |
| KM | Transferase; Glycosyltransferase; Glycoprotein; Transmembrane; signal, |
| KM | MultiGene family; Microsome; Polysialylation. |
| FT | SIGNAL |
| FT | CHAIN |
| FT | TRANSMEM |
| FT | CARBOHYD |
| VARIANT | |
| FT | VARIANT |
| FT | VARIANT |
| FT | VARIANT |
| FT | VARIANT |
| FT | VARIANT |
| FT | CONFLICT |
| FT | CONFLICT |
| FT | CONFLICT |
| SO | SEQUENCE |

Matches 492; Conservative 14; Mismatches 22; Indels 1; Gaps 1

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
84

QY 61 SISFDPNPSTLKFEVYPVSLTKTEFEDI IKOLVKRWAE LPKDTEWSESVSOVQETMMTEN 120

61 SISFDPNPSTLKFEVYPVSLTKTEFEDI IKQLVKRWAE LPKDTFWSYFSQVQEIMWTFN 120

[illegible]

181 EKHSGGLLFPSPYVPVMSEI.SDOMTFIRVKNMITYVT.YEEFEWOTEDMKKWDQVCEIIT 240

1. The first step is to identify the problem or question that needs to be answered. This involves understanding the context and the specific requirements of the task.

211 GAFILDEINANDIMLIKINWDFQFPHPLPNVEFVGLHCKPAKPLPKEMEETFVQSSG 300

201 ENCIPHERED COMMUNICATIONS 111

| | | | | | | |
|----|---------|-------|-----------|------|-----|-----|
| ID | _UDBJ | MACFA | STANDARD; | PRT; | 528 | AA. |
| AC | 09XT55. | | | | | |

DI 30-MAY-2000 (Rel. 39, last sequence update)
DT 16-OCT-2001 (Rel. 40, last annotation update)

05 *Macaca fascicularis* (Crab eating macaque) (*Cynomolgus* monkey).
06 Enkavota, Meratza, Chordata, Craniata, Vertebrata, Mammalia,
07 Eukaryota.

CC Cercopithecinae; Macaca.
OX NCBI_TaxID=9541;

RA Belanger G., Barbier O., Hum D.W., Belanger A.;

RT testosterone,";
RL Eur. J. Biochem. 260:701-708 (1999).

CC CONSEQUENT ELIMINATION OF POTENTIAL BIOALKALINOBIOLOGICS AND
CC ENDOGENOUS COMPOUNDS. THIS ISOZYME DISPLAYS ACTIVITY TOWARD
CC ACTIVITY OF 100% ON VARIOUS SUBSTRATES.

CC NARLINGENIN AND O,O'-BIPHENOL, ACTIVE ALSO ON 3A-HYDROXY AND 17B-
CC HYDROXY POSITIONS OF STEROIDS.

- CATALYTIC ACTIVITY: UDP-glucuronate + acceptor = UDP + acceptor

CC - TISSUE SPECIFICITY: EXPRESSED IN LIVER, OVARY, PROSTATE, COLON, KIDNEY, PANCREAS, BRAIN, CEREBELLUM, MAMMARY GLAND AND EPIDIDYMIS

CC -I- SIMILARITY: BELONGS TO THE UDP-GLYCOSYLTRANSFERASE FAMILY.

CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on it
CC

entitles requires a license agreement (See <http://www.isb-sib.ch/announce>)

DR EMBL; AF112112; AAD24435.1; -.
DR InterPro; IPR002213; UDP_gluco trans.

Transferrase; Glycosyltransferase; Glycoprotein; Signal; Membrane protein; Transmembrane; Signal; Membrane protein

| | | | | |
|----|----------|-----|-----|-----------------------------------|
| FT | CHAIN | 22 | 528 | UDP-GLUCURONOSYLTRANSFERASE 2B19. |
| FT | TRANSMEM | 493 | 513 | POTENTIAL. |

[illegible]

THIS PAGE BLANK (USPTO)

841 ACTCTTACCAATGTTGAGTTCGTTGGAGACTCCACTGCAAACTGCAAAACCCCTACC 9

JOURNAL Biochem. J. 242 (2), 581-588 (1987)
MEDLINE 87241362
PUBMED 3109396

MEDLINE 87241362
PUBMED 3109396

THIS PAGE BLANK (USPTO)

THIS PAGE BLANK (USPTO)

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 5, 2003, 09:47:25 ; Search time 21 Seconds
(without alignments)
1065.631 Million cell updates/sec

Title: US-09-980-729B-5

Perfect score: 2798
Sequence: 1 MSKMTSALLIOLSCYFSS.....KCLFSCQKFKTRKIERRE 529

Scoring table: BLOSUM62
Gapop 10.0, Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database:

Issued Patents AA:*
1: /cgn2_6/ptodata/1/1aa/5A COMB.pep.*
2: /cgn2_6/ptodata/1/1aa/5B COMB.pep.*
3: /cgn2_6/ptodata/1/1aa/6A COMB.pep.*
4: /cgn2_6/ptodata/1/1aa/6B COMB.pep.*
5: /cgn2_6/ptodata/1/1aa/PTUS COMB.pep.*
6: /cgn2_6/ptodata/1/1aa/backfile1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|--------|-------------|--------|----|-------------------|
| 1 | 2602.5 | 93.0 | 528 | 4 | US-09-356-806-8 |
| 2 | 2322 | 83.0 | 524 | 4 | US-09-356-806-40 |
| 3 | 2169.5 | 77.5 | 530 | 3 | US-09-180-852-2 |
| 4 | 2158.5 | 77.1 | 530 | 4 | US-09-356-806-113 |
| 5 | 1874.5 | 67.0 | 454 | 4 | US-09-813-918-2 |
| 6 | 1243 | 44.4 | 288 | 4 | US-09-813-918-3 |
| 7 | 1196.5 | 42.8 | 531 | 5 | PCT-US92-00282-6 |
| 8 | 1130.5 | 40.4 | 533 | 5 | PCT-US92-00282-3 |
| 9 | 1126 | 40.2 | 534 | 5 | PCT-US92-00282-4 |
| 10 | 1064 | 38.0 | 531 | 5 | PCT-US92-00282-5 |
| 11 | 1045.5 | 37.4 | 529 | 5 | PCT-US92-00282-7 |
| 12 | 833 | 29.8 | 197 | 4 | US-09-813-918-4 |
| 13 | 765.5 | 27.4 | 245 | 4 | US-09-305-856B-18 |
| 14 | 403 | 14.4 | 515 | 3 | US-08-942-012B-32 |
| 15 | 393 | 14.0 | 129 | 4 | US-09-370-838-36 |
| 16 | 379 | 13.5 | 488 | 3 | US-08-942-012B-29 |
| 17 | 379 | 13.5 | 488 | 3 | US-08-942-012B-30 |
| 18 | 379 | 13.5 | 488 | 3 | US-08-942-012B-33 |
| 19 | 372 | 13.3 | 288 | 4 | US-09-305-856B-2 |
| 20 | 372 | 13.3 | 288 | 4 | PCT-US92-00282-19 |
| 21 | 362.5 | 13.0 | 289 | 4 | US-09-305-856B-6 |
| 22 | 360.5 | 12.9 | 289 | 4 | US-09-305-856B-8 |
| 23 | 360.5 | 12.9 | 289 | 4 | PCT-US92-00282-13 |
| 24 | 358.5 | 12.8 | 289 | 4 | US-09-305-856B-4 |
| 25 | 358.5 | 12.8 | 289 | 4 | PCT-US92-00282-15 |
| 26 | 354.5 | 12.7 | 289 | 5 | PCT-US92-00282-11 |
| 27 | 346 | 12.4 | 310 | 4 | US-09-305-856B-14 |

| | | | | | | |
|----|-------|------|-----|---|-------------------|--------------------|
| 28 | 338 | 12.1 | 515 | 3 | US-08-942-01 | Sequence 24, Appl |
| 29 | 321.5 | 11.5 | 317 | 4 | US-09-305-856B-10 | Sequence 12, Appl |
| 30 | 316 | 11.3 | 287 | 4 | US-09-305-856B-10 | Sequence 10, Appl |
| 31 | 306.5 | 11.0 | 286 | 5 | PCT-US92-00282-9 | Sequence 9, Appl |
| 32 | 297.5 | 10.6 | 98 | 5 | PCT-US92-00282-26 | Sequence 25, Appl |
| 33 | 296.5 | 10.6 | 253 | 4 | US-09-305-856B-16 | Sequence 16, Appl |
| 34 | 294.5 | 10.5 | 506 | 3 | US-08-942-012B-26 | Sequence 25, Appl |
| 35 | 287 | 10.3 | 506 | 3 | US-08-942-012B-25 | Sequence 25, Appl |
| 36 | 287 | 10.3 | 506 | 6 | 5180581-2 | Patent No. 5180581 |
| 37 | 271 | 9.7 | 58 | 2 | US-08-466-583-9 | Sequence 9, Appl |
| 38 | 271 | 9.7 | 58 | 4 | US-08-265-427-9 | Sequence 9, Appl |
| 39 | 271 | 9.7 | 58 | 5 | PCT-US95-07820-9 | Sequence 9, Appl |
| 40 | 261 | 9.3 | 493 | 3 | US-08-942-012B-28 | Sequence 28, Appl |
| 41 | 257 | 9.2 | 52 | 2 | US-08-466-583-7 | Sequence 7, Appl |
| 42 | 257 | 9.2 | 52 | 4 | US-08-265-427-7 | Sequence 7, Appl |
| 43 | 257 | 9.2 | 52 | 4 | PCT-US95-07820-7 | Sequence 7, Appl |
| 44 | 250.5 | 9.0 | 489 | 3 | US-08-942-012B-31 | Sequence 31, Appl |
| 45 | 241.5 | 8.6 | 491 | 3 | US-08-942-012B-27 | Sequence 27, Appl |

ALIGNMENTS

RESULT 1
US-09-356-806-8
Sequence 8, Application US/09356806

Patent No. 6586175
GENERAL INFORMATION:
APPLICANT: Penny, Laura
APPLICANT: Galvin, Margaret
APPLICANT: Miller, Andrew
APPLICANT: Reidy, Michael
TITLE OF INVENTION: Genotyping Human
TITLE OF INVENTION: UDP-Glucuronosyltransferase 2B4 (UGT2B4), 2B7 (UGT2B7) and
TITLE OF INVENTION: 2B15 (UGT2B15) Genes
FILE REFERENCE: SEQ-22PRV2
CURRENT APPLICATION NUMBER: US/09/356,806
CURRENT FILING DATE: 1999-07-20
NUMBER OF SEQ ID NOS: 164
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 8
LENGTH: 528
TYPE: PRT
ORGANISM: H. sapiens
US-09-356-806-8

Query Match 93.0%; Score 2602.5; DB 4; Length 528;
Best Local Similarity 93.0%; Pred. No. 4.7e-263;
Matches 492; Conservative 14; Mismatches 22; Indels 1; Gaps 1;

| | | |
|----|-----|---|
| QY | 1 | MSKMTSALLIOLSCYFSSGCGKYLWPTFESHMNNIKTIIDELVORGHETVLA5SA 60 |
| DB | 1 | MSKMTSALLIOLSCYFSSGCGKYLWPTFESHMNNIKTIIDELVORGHETVLA5SA 60 |
| QY | 61 | SISDPGSPSTLKEVYVPLTTEFEDIKOLVKRAELPKDTFMSYFQVOEIMWTFN 120 |
| DB | 61 | SISDPGSPSTLKEVYVPLTTEFEDIKOLVKRAELPKDTFMSYFQVOEIMWTFN 120 |
| QY | 121 | DIILKPKDVISNKKMKKLQESREEDVTLADAVFPFELLAEELKIFPVYSLRSPGAI 180 |
| DB | 121 | DIILKPKDVISNKKMKKLQESREEDVTLADAVFPFELLAEELKIFPVYSLRSPGAI 180 |
| QY | 181 | EKHSGLLFPSPYVYVMSSELSDMTIERVKNNIYLYFEFVQIDMKMDQFYSVL 240 |
| DB | 181 | EKHSGLLFPSPYVYVMSSELSDMTIERVKNNIYLYFEFVQIDMKMDQFYSVL 240 |
| QY | 241 | GRPTTISPMKADIMILIRNYWDFQPHPLPNVEFVGLCKCPAKPLPKMEEFVQSSG 300 |
| DB | 241 | GRPTTISPMKADIMILIRNYWDFQPHPLPNVEFVGLCKCPAKPLPKMEEFVQSSG 300 |
| QY | 301 | ENGWVVFSLGMSVNTSEERANVTASALAKIPOKYLWFPDGNKPDTLGINTRLKWTIPON 360 |
| DB | 301 | ENGWVVFSLGMSVNTSEERANVTASALAKIPOKYLWFPDGNKPDTLGINTRLKWTIPON 360 |

THIS PAGE BLANK (USP10)

| Query Match | Similarity | 83.0% | Score 2302 | DB 4: | Length 524 |
|-------------|--------------|-------|--------------------|-------|------------|
| Best Local | Similarity | 81.5% | Pred. No. 9.7e-234 | | |
| Matches | Conservative | 427 | Mismatches | 35 | Indels |
| | | | | | Gaps |
| | | | | | 0 |

| Query | Subject | Score | Length | Identical | Similarity |
|-------|--|-------|--------|-----------|------------|
| Q1 | MSMKVTSALLLITLSCYFSSGSCGVLVLPTEFSHWNKITLDELVQGHETVLASSA | 60 | 60 | 100 | 100 |
| Db | 1 MSVKVTSVILLITQLSFCESSNGCGLVWMAEYSHMNIKITLDELQGHETVLASSA | 60 | 60 | 100 | 100 |
| Q2 | 61 SLSGPNBSBTLKFEVYVSLTKTEFEDIIKOLVYKMAELPKOTFWSYFSQVQEIWNTN | 120 | 120 | 100 | 100 |
| Db | 61 SLPFPNNSSALKIETIYPTSLTKTELENIWQIKRMSDLPKOTFWLYFSQVQEIWSING | 120 | 120 | 100 | 100 |
| Q3 | 121 DILRRCNDIVNKKLMMKKLOESRPDVLADAVPFGGELLAEILKLPFYSLRFSFGYAI | 180 | 180 | 100 | 100 |
| Db | 121 DITRFCKCDVSNKKFMKKVQESRPDVPADAIIPCESELAELFNLPFYSLRFSFGYAF | 180 | 180 | 100 | 100 |
| Q4 | 181 EKHSGGLFPPEYVYVNMSELSDQNTFIRVKNMIVYLFEFEPFOIFDMKKMDQFSEVL | 240 | 240 | 100 | 100 |
| Db | 181 EKHSGGFLFPPEYVYVNMSELTDQNTFERVKNMIVYLFDFMEFLFDMKKMDQFSEVL | 240 | 240 | 100 | 100 |
| Q5 | 241 GRPTLLSETMAKADIMLIRNYWDFQFPHDLINVEVFGLHCKCPAKPLPKEMEEFPQSSG | 300 | 300 | 100 | 100 |
| Db | 241 GRPTLLSETMGADVWLIRNSMNPQFPRYLLNVDFVGGHLCKCPAKPLPKEMEDFPQSSG | 300 | 300 | 100 | 100 |
| Q6 | 301 ENGVVVFSLGSVNSVSTSEERANVASALAKIPQKVLMPFDGKKPDTLLGNTLRKYKIPGN | 360 | 360 | 100 | 100 |
| Db | 301 ENGVVVFSLGSVNSVSTSEERANVASALAQIPQKVLMPFDGKKPDTLLGNTLRKYKIPGN | 360 | 360 | 100 | 100 |
| Q7 | 361 DILGHPKTKAFLTHGGMNGIYEAITHGVNMGVPIFGDLDNIIAHKKAGAAVEINFKM | 420 | 420 | 100 | 100 |
| Db | 361 DILGHPKTKAFLTHGGMNGIYEAITHGVNIGVPIFGDLDNIIAHKKAGAAVVDPNM | 420 | 420 | 100 | 100 |
| Q8 | 421 TSEDILRLARTVITQSSYKENAMRLSRIHDDPVPKLDAAVMIIEFVMMHKAKHLRSA | 480 | 480 | 100 | 100 |
| Db | 421 SSTDILNLAKVIYNDPSYKENAMKLSRIHDDPVPKLDAAVMIIEFVMMHKAKHLRVA | 480 | 480 | 100 | 100 |

THIS PAGE BLANK (USRIU)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 5, 2003, 09:49:45 ; Search time 36 Seconds

(without alignments)
2732.927 Million cell updates/sec

Title: US-09-980-729B-5

Perfect score: 2798

Sequence: 1 MSMKWTALLLIOLSCYFSS.....KCLFSCQCFKTKRIEKR 529

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 684280 seqs, 185983659 residues

Total number of hits satisfying chosen parameters: 684280

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

Published Applications_Aa.*
1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep.*
7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep.*
10: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep.*
11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep.*
12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep.*
13: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB.pep.*
14: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB.pep.*
15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep.*
16: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep.*
17: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*
18: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|--------|-------------|--------|-------------------|-------------------|
| 1 | 2602.5 | 93.0 | 528 | US-10-205-522-8 | Sequence 8, Appl1 |
| 2 | 2330 | 83.3 | 529 | US-09-981-353-194 | Sequence 194, App |
| 3 | 2330 | 83.3 | 529 | US-10-057-834A-2 | Sequence 2, Appl1 |
| 4 | 2322 | 83.0 | 524 | US-10-205-522-40 | Sequence 40, Appl |
| 5 | 2158.5 | 77.1 | 530 | US-10-205-522-113 | Sequence 113, App |
| 6 | 1924.5 | 68.8 | 527 | US-09-981-353-166 | Sequence 166, App |
| 7 | 1924.5 | 68.8 | 527 | US-10-199-672-522 | Sequence 522, App |
| 8 | 1924.5 | 68.8 | 527 | US-10-187-749-522 | Sequence 522, App |
| 9 | 1924.5 | 68.8 | 527 | US-10-194-457-522 | Sequence 522, App |
| 10 | 1924.5 | 68.8 | 527 | US-10-184-642-522 | Sequence 522, App |
| 11 | 1924.5 | 68.8 | 527 | US-10-196-747-522 | Sequence 522, App |
| 12 | 1924.5 | 68.8 | 527 | US-10-173-689-522 | Sequence 522, App |
| 13 | 1924.5 | 68.8 | 527 | US-10-173-689-522 | Sequence 522, App |
| 14 | 1924.5 | 68.8 | 527 | US-10-173-691-522 | Sequence 522, App |
| 15 | 1924.5 | 68.8 | 527 | US-10-173-692-522 | Sequence 522, App |

| | | | | | |
|----|--------|------|-----|-------------------|-------------------|
| 16 | 1924.5 | 68.8 | 527 | US-10-173-694-522 | Sequence 522, App |
| 17 | 1924.5 | 68.8 | 527 | US-10-173-698-522 | Sequence 522, App |
| 18 | 1924.5 | 68.8 | 527 | US-10-173-699-522 | Sequence 522, App |
| 19 | 1924.5 | 68.8 | 527 | US-10-173-707-522 | Sequence 522, App |
| 20 | 1924.5 | 68.8 | 527 | US-10-174-569-522 | Sequence 522, App |
| 21 | 1924.5 | 68.8 | 527 | US-10-174-583-522 | Sequence 522, App |
| 22 | 1924.5 | 68.8 | 527 | US-10-174-587-522 | Sequence 522, App |
| 23 | 1924.5 | 68.8 | 527 | US-10-174-589-522 | Sequence 522, App |
| 24 | 1924.5 | 68.8 | 527 | US-10-175-591-522 | Sequence 522, App |
| 25 | 1924.5 | 68.8 | 527 | US-10-175-736-522 | Sequence 522, App |
| 26 | 1924.5 | 68.8 | 527 | US-10-175-742-522 | Sequence 522, App |
| 27 | 1924.5 | 68.8 | 527 | US-10-175-745-522 | Sequence 522, App |
| 28 | 1924.5 | 68.8 | 527 | US-10-175-745-522 | Sequence 522, App |
| 29 | 1924.5 | 68.8 | 527 | US-10-175-751-522 | Sequence 522, App |
| 30 | 1924.5 | 68.8 | 527 | US-10-175-751-522 | Sequence 522, App |
| 31 | 1924.5 | 68.8 | 527 | US-10-175-754-522 | Sequence 522, App |
| 32 | 1924.5 | 68.8 | 527 | US-10-176-480-522 | Sequence 522, App |
| 33 | 1924.5 | 68.8 | 527 | US-10-176-489-522 | Sequence 522, App |
| 34 | 1924.5 | 68.8 | 527 | US-10-176-754-522 | Sequence 522, App |
| 35 | 1924.5 | 68.8 | 527 | US-10-176-755-522 | Sequence 522, App |
| 36 | 1924.5 | 68.8 | 527 | US-10-176-759-522 | Sequence 522, App |
| 37 | 1924.5 | 68.8 | 527 | US-10-176-920-522 | Sequence 522, App |
| 38 | 1924.5 | 68.8 | 527 | US-10-176-924-522 | Sequence 522, App |
| 39 | 1924.5 | 68.8 | 527 | US-10-176-924-522 | Sequence 522, App |
| 40 | 1924.5 | 68.8 | 527 | US-10-176-984-522 | Sequence 522, App |
| 41 | 1924.5 | 68.8 | 527 | US-10-179-508-522 | Sequence 522, App |
| 42 | 1924.5 | 68.8 | 527 | US-10-179-512-522 | Sequence 522, App |
| 43 | 1924.5 | 68.8 | 527 | US-10-179-515-522 | Sequence 522, App |
| 44 | 1924.5 | 68.8 | 527 | US-10-173-702-522 | Sequence 522, App |
| 45 | 1924.5 | 68.8 | 527 | US-10-173-703-522 | Sequence 522, App |

ALIGNMENTS

RESULT 1
US-10-205-522-8
Sequence 8 Application US/10205522
Publication No. US20030077629A1
GENERAL INFORMATION:
APPLICANT: Penny, Laura
APPLICANT: Galvin, Margaret
APPLICANT: Miller, Andrew
APPLICANT: Reidy, Michael
TITLE OF INVENTION: Genotyping Human
TITLE OF INVENTION: UDP-Glucuronosyltransferase 2B4 (UGT2B4), 2B7 (UGT2B7) and
FILE REFERENCE: SEQ-22PRV2
CURRENT FILING DATE: 2002-07-24
PRIORITY FILING DATE: 1999-07-20
PRIORITY FILING DATE: 1999-07-20
NUMBER OF SEQ ID NOS: 164
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 8
LENGTH: 528
TYPE: PRT
ORGANISM: H. sapiens
US-10-205-522-8
Query Match 93.0%; Score 2602.5; DB 15; Length 528;
Best Local Similarity 93.0%; Pred. No. 8.5e-53;
Matches 492; Conservative 14; Mismatches 22; Indels 1; Gaps 1;
QY 1 MSMKWTALLLIOLSCYFSSGCGKVLWPTFESHMNNIKITIDELVORGEVTLASSA 60
Db 1 MSMKWTALLLIOLSCYFSSGCGKVLWPTFESHMNNIKITIDELVORGEVTLASSA 60
QY 61 SISFDPSPELTKEFYVPSVLTETFEEDITKOLVKMAELPKOTFMSYFQVOEIMWTFN 120
Db 61 SISFDPSPELTKEFYVPSVLTETFEEDITKOLVKMAELPKOTFMSYFQVOEIMWTFN 120
QY 121 DIRKFKDIVSKKLMKQLQESRFVDVLAIVFPFGELLAEELKIPFYVSLFSPGYAI 180

THIS PAGE BLANK (USPTO)

```

Db 121 DILRFCKDQIVSNKKLMKKLOESRPDVVLADAVFPFGLLAEKLPVYSLRFSFGAI 180
Qy 181 EKHSGGLFPSPSYVAVVMSLSDQMTFIERVKNMIVYLFEFPOIPMKKKMDQYSEVL 240
Db 181 EKHSGGLFPSPSYVAVVMSLSDQMTFIERVKNMIVYLFEFPOIPMKKKMDQYSEVL 240
Qy 241 GRPPTLSETMAKADIMLRNYWDFOPHPLPNVEFVGLHCKPAKPLPKEMEFPVQSSG 300
Db 241 GRPPTLSETMAKADIMLRNYWDFOPHPLPNVEFVGLHCKPAKPLPKEMEFPVQSSG 300
Qy 301 ENGVAVFSLGSMVSNTESEERANVIASALAKIPQKVLMPFDGKPDGLNTRLYKMIPON 360
Db 301 ENGVAVFSLGSMVSNTESEERANVIASALAKIPQKVLMPFDGKPDGLNTRLYKMIPON 360
Qy 361 DLGHPKTKAFITHGANGIYEAIYHGVPMVGPPIFGDQDNIAMHKAAGAVEINFKTM 420
Db 361 DLGHPKTKAFITHGANGIYEAIYHGVPMVGPPIFGDQDNIAMHKAAGAVEINFKTM 420
Qy 421 TSIEDLLRALRTVITDSSYKENAMRLSRHHDPVKPLDRAVFWIEFVMRHKGAKHLRSAA 480
Db 421 TSIEDLLRALRTVITDSSYKENAMRLSRHHDPVKPLDRAVFWIEFVMRHKGAKHLRSAA 480
Qy 481 HDLTFPOHYSIDVIGFLITCVATAIFLFTKCFLFCQCKFNKTRKIEKRE 529
Db 481 HDLTFPOHYSIDVIGFLITCVATAIFLFTKCFLFCQCKFNKTRKIEKRE 529

```

RESULT 2

```

US-09-981-353-194
; Sequence 194, Application US/09981353
; Patent No. US20020160382A1
; GENERAL INFORMATION:
; APPLICANT: Lasek, Amy W.
; APPLICANT: Jones, David A.
; TITLE OF INVENTION: GENES EXPRESSED IN COLON CANCER
; FILE REFERENCE: PA-0038 US
; CURRENT APPLICATION NUMBER: US/09/981,353
; CURRENT FILING DATE: 2001-10-11
; NUMBER OF SEQ ID NOS: 194
; SOFTWARE: PERL Program
; SEQ ID NO: 194
; LENGTH: 529
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. US20020160382A1 088078CD1
US-09-981-353-194

```

Query Match

```

Best Local Similarity 81.1%; Score 2330; DB 10; Length 529;
Matches 429; Conservative 42; Mismatches 58; Indels 0; Gaps 0;

```

```

Qy 1 MSKMTSALLLIQLSCYFSSGCGKVLVPTFESHMNIKITLDELVORGEVTVLASSA 60
Db 1 MSKMTSALLLIQLSCYFSSGCGKVLVPTFESHMNIKITLDELVORGEVTVLASSA 60
Qy 61 SISDPNPSPTLKPEYVPVSLTKTEPEDIIKOLVKRMAELPKDTWMSYFSQVOEIMMTFN 120
Db 61 SISDPNPSPTLKPEYVPVSLTKTEPEDIIKOLVKRMAELPKDTWMSYFSQVOEIMMTFN 120
Qy 121 DILRFCKDQIVSNKKLMKKLOESRPDVVLADAVFPFGLLAEKLPVYSLRFSFGAI 180
Db 121 DILRFCKDQIVSNKKLMKKLOESRPDVVLADAVFPFGLLAEKLPVYSLRFSFGAI 180
Qy 181 EKHSGGLFPSPSYVAVVMSLSDQMTFIERVKNMIVYLFEFPOIPMKKKMDQYSEVL 240
Db 181 EKHSGGLFPSPSYVAVVMSLSDQMTFIERVKNMIVYLFEFPOIPMKKKMDQYSEVL 240
Qy 241 GRPPTLSETMAKADIMLRNYWDFOPHPLPNVEFVGLHCKPAKPLPKEMEFPVQSSG 300
Db 241 GRPPTLSETMAKADIMLRNYWDFOPHPLPNVEFVGLHCKPAKPLPKEMEFPVQSSG 300

```

```

Qy 301 ENGVAVFSLGSMVSNTESEERANVIASALAKIPQKVLMPFDGKPDGLNTRLYKMIPON 360
Db 301 ENGVAVFSLGSMVSNTESEERANVIASALAKIPQKVLMPFDGKPDGLNTRLYKMIPON 360
Qy 361 DLGHPKTKAFITHGANGIYEAIYHGVPMVGPPIFGDQDNIAMHKAAGAVEINFKTM 420
Db 361 DLGHPKTKAFITHGANGIYEAIYHGVPMVGPPIFGDQDNIAMHKAAGAVEINFKTM 420
Qy 421 TSIEDLLRALRTVITDSSYKENAMRLSRHHDPVKPLDRAVFWIEFVMRHKGAKHLRSAA 480
Db 421 TSIEDLLRALRTVITDSSYKENAMRLSRHHDPVKPLDRAVFWIEFVMRHKGAKHLRSAA 480
Qy 481 HDLTFPOHYSIDVIGFLITCVATAIFLFTKCFLFCQCKFNKTRKIEKRE 529
Db 481 HDLTFPOHYSIDVIGFLITCVATAIFLFTKCFLFCQCKFNKTRKIEKRE 529

```

RESULT 3

```

US-10-057-834A-2
; Sequence 2, Application US/10057834A
; Publication No. US2003009960A1
; GENERAL INFORMATION:
; APPLICANT: RATTIN, MARK J.
; APPLICANT: INNOCENTI, FEDERICO
; APPLICANT: IYER, LALITHA
; APPLICANT: DAS, SOMA
; APPLICANT: SAWYER, MICHAEL
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR OPTIMIZING UGT2B7 SUBSTRATE DOSINGS
; FILE REFERENCE: ARCD:358US
; CURRENT APPLICATION NUMBER: US/10/057,834A
; CURRENT FILING DATE: 2002-08-22
; PRIOR APPLICATION NUMBER: UNKNOWN
; PRIOR FILING DATE: 2002-01-25
; NUMBER OF SEQ ID NOS: 78
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO: 2
; LENGTH: 529
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-057-834A-2

```

```

Query Match 83.3%; Score 2330; DB 15; Length 529;
Best Local Similarity 81.1%; Pred. No. 2.3e-225;
Matches 429; Conservative 42; Mismatches 58; Indels 0; Gaps 0;

```

```

Qy 1 MSKMTSALLLIQLSCYFSSGCGKVLVPTFESHMNIKITLDELVORGEVTVLASSA 60
Db 1 MSKMTSALLLIQLSCYFSSGCGKVLVPTFESHMNIKITLDELVORGEVTVLASSA 60
Qy 61 SISDPNPSPTLKPEYVPVSLTKTEPEDIIKOLVKRMAELPKDTWMSYFSQVOEIMMTFN 120
Db 61 SISDPNPSPTLKPEYVPVSLTKTEPEDIIKOLVKRMAELPKDTWMSYFSQVOEIMMTFN 120
Qy 121 DILRFCKDQIVSNKKLMKKLOESRPDVVLADAVFPFGLLAEKLPVYSLRFSFGAI 180
Db 121 DILRFCKDQIVSNKKLMKKLOESRPDVVLADAVFPFGLLAEKLPVYSLRFSFGAI 180
Qy 181 EKHSGGLFPSPSYVAVVMSLSDQMTFIERVKNMIVYLFEFPOIPMKKKMDQYSEVL 240
Db 181 EKHSGGLFPSPSYVAVVMSLSDQMTFIERVKNMIVYLFEFPOIPMKKKMDQYSEVL 240
Qy 241 GRPPTLSETMAKADIMLRNYWDFOPHPLPNVEFVGLHCKPAKPLPKEMEFPVQSSG 300
Db 241 GRPPTLSETMAKADIMLRNYWDFOPHPLPNVEFVGLHCKPAKPLPKEMEFPVQSSG 300
Qy 301 ENGVAVFSLGSMVSNTESEERANVIASALAKIPQKVLMPFDGKPDGLNTRLYKMIPON 360
Db 301 ENGVAVFSLGSMVSNTESEERANVIASALAKIPQKVLMPFDGKPDGLNTRLYKMIPON 360
Qy 361 DLGHPKTKAFITHGANGIYEAIYHGVPMVGPPIFGDQDNIAMHKAAGAVEINFKTM 420
Db 361 DLGHPKTKAFITHGANGIYEAIYHGVPMVGPPIFGDQDNIAMHKAAGAVEINFKTM 420

```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 5, 2003, 09:49:45 ; Search time 36 Seconds
(without alignments)
2732.927 Million cell updates/sec

Title: US-09-980-729b-5
Perfect score: 2798
Sequence: 1 MSKMTSALLLIQUSCYFSS.....KCFLFSCQKFKTKIEKRE 529

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 684280 seqs, 185983659 residues

Total number of hits satisfying chosen parameters: 684280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

Published Applications NA:
1: /cgn2_6/ptodata/1/pubppa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/1/pubppa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/1/pubppa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/1/pubppa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/1/pubppa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/1/pubppa/PCTUS_PUBCOMB.pep.*
7: /cgn2_6/ptodata/1/pubppa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/1/pubppa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/1/pubppa/US09_PUBCOMB.pep.*
10: /cgn2_6/ptodata/1/pubppa/US09B_PUBCOMB.pep.*
11: /cgn2_6/ptodata/1/pubppa/US09C_PUBCOMB.pep.*
12: /cgn2_6/ptodata/1/pubppa/US09_NEW_PUB.pep.*
13: /cgn2_6/ptodata/1/pubppa/US10A_PUBCOMB.pep.*
14: /cgn2_6/ptodata/1/pubppa/US10B_PUBCOMB.pep.*
15: /cgn2_6/ptodata/1/pubppa/US10C_PUBCOMB.pep.*
16: /cgn2_6/ptodata/1/pubppa/US10_NEW_PUB.pep.*
17: /cgn2_6/ptodata/1/pubppa/US60_NEW_PUB.pep.*
18: /cgn2_6/ptodata/1/pubppa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|--------|-------------|--------|----|-------------------|
| 1 | 2602.5 | 93.0 | 528 | 15 | US-10-205-522-8 |
| 2 | 2330 | 83.3 | 529 | 10 | US-09-981-353-194 |
| 3 | 2330 | 83.3 | 529 | 15 | US-10-057-834A-2 |
| 4 | 2322 | 83.0 | 524 | 15 | US-10-205-522-40 |
| 5 | 2158.5 | 77.1 | 530 | 15 | US-10-205-522-113 |
| 6 | 1924.5 | 68.8 | 527 | 10 | US-09-981-353-166 |
| 7 | 1924.5 | 68.8 | 527 | 12 | US-10-199-672-522 |
| 8 | 1924.5 | 68.8 | 527 | 12 | US-10-187-749-522 |
| 9 | 1924.5 | 68.8 | 527 | 12 | US-10-194-457-522 |
| 10 | 1924.5 | 68.8 | 527 | 12 | US-10-184-642-522 |
| 11 | 1924.5 | 68.8 | 527 | 12 | US-10-196-747-522 |
| 12 | 1924.5 | 68.8 | 527 | 12 | US-10-173-689-522 |
| 13 | 1924.5 | 68.8 | 527 | 12 | US-10-173-690-522 |
| 14 | 1924.5 | 68.8 | 527 | 12 | US-10-173-691-522 |
| 15 | 1924.5 | 68.8 | 527 | 12 | US-10-173-692-522 |

| | | | | | | |
|----|--------|------|-----|----|-------------------|-------------------|
| 16 | 1924.5 | 68.8 | 527 | 12 | US-10-173-694-522 | Sequence 522, App |
| 17 | 1924.5 | 68.8 | 527 | 12 | US-10-173-698-522 | Sequence 522, App |
| 18 | 1924.5 | 68.8 | 527 | 12 | US-10-173-699-522 | Sequence 522, App |
| 19 | 1924.5 | 68.8 | 527 | 12 | US-10-173-707-522 | Sequence 522, App |
| 20 | 1924.5 | 68.8 | 527 | 12 | US-10-174-569-522 | Sequence 522, App |
| 21 | 1924.5 | 68.8 | 527 | 12 | US-10-174-583-522 | Sequence 522, App |
| 22 | 1924.5 | 68.8 | 527 | 12 | US-10-174-587-522 | Sequence 522, App |
| 23 | 1924.5 | 68.8 | 527 | 12 | US-10-174-589-522 | Sequence 522, App |
| 24 | 1924.5 | 68.8 | 527 | 12 | US-10-174-591-522 | Sequence 522, App |
| 25 | 1924.5 | 68.8 | 527 | 12 | US-10-175-736-522 | Sequence 522, App |
| 26 | 1924.5 | 68.8 | 527 | 12 | US-10-175-742-522 | Sequence 522, App |
| 27 | 1924.5 | 68.8 | 527 | 12 | US-10-175-745-522 | Sequence 522, App |
| 28 | 1924.5 | 68.8 | 527 | 12 | US-10-175-745-522 | Sequence 522, App |
| 29 | 1924.5 | 68.8 | 527 | 12 | US-10-175-748-522 | Sequence 522, App |
| 30 | 1924.5 | 68.8 | 527 | 12 | US-10-175-751-522 | Sequence 522, App |
| 31 | 1924.5 | 68.8 | 527 | 12 | US-10-175-754-522 | Sequence 522, App |
| 32 | 1924.5 | 68.8 | 527 | 12 | US-10-176-480-522 | Sequence 522, App |
| 33 | 1924.5 | 68.8 | 527 | 12 | US-10-176-489-522 | Sequence 522, App |
| 34 | 1924.5 | 68.8 | 527 | 12 | US-10-176-754-522 | Sequence 522, App |
| 35 | 1924.5 | 68.8 | 527 | 12 | US-10-176-755-522 | Sequence 522, App |
| 36 | 1924.5 | 68.8 | 527 | 12 | US-10-176-759-522 | Sequence 522, App |
| 37 | 1924.5 | 68.8 | 527 | 12 | US-10-176-920-522 | Sequence 522, App |
| 38 | 1924.5 | 68.8 | 527 | 12 | US-10-176-922-522 | Sequence 522, App |
| 39 | 1924.5 | 68.8 | 527 | 12 | US-10-176-924-522 | Sequence 522, App |
| 40 | 1924.5 | 68.8 | 527 | 12 | US-10-176-984-522 | Sequence 522, App |
| 41 | 1924.5 | 68.8 | 527 | 12 | US-10-179-508-522 | Sequence 522, App |
| 42 | 1924.5 | 68.8 | 527 | 12 | US-10-179-512-522 | Sequence 522, App |
| 43 | 1924.5 | 68.8 | 527 | 12 | US-10-179-515-522 | Sequence 522, App |
| 44 | 1924.5 | 68.8 | 527 | 12 | US-10-173-702-522 | Sequence 522, App |
| 45 | 1924.5 | 68.8 | 527 | 12 | US-10-173-703-522 | Sequence 522, App |

ALIGNMENTS

RESULT 1
US-10-205-522-8
Sequence-8: Application US/10205522
Publication No. US20030077629A1
GENERAL INFORMATION:
APPLICANT: Penny, Laura
APPLICANT: Galvin, Margaret
APPLICANT: Miller, Andrew
APPLICANT: Reidy, Michael
TITLE OF INVENTION: Genotyping Human
TITLE OF INVENTION: UDP-Glucuronosyltransferase 2B4 (UGT2B4), 2B7 (UGT2B7) and
FILE REFERENCE: SEQ-22PRV2
CURRENT FILING DATE: US/10/205,522
PRIORITY FILING DATE: 2002-07-24
PRIOR APPLICATION NUMBER: US/09/356,806
PRIOR FILING DATE: 1999-07-20
NUMBER OF SEQ ID NOS: 164
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 8
LENGTH: 528
TYPE: PRT
ORGANISM: H. sapiens
US-10-205-522-8
Query Match 93.0%; Score 2602.5; DB 15; Length 528;
Best Local Similarity 93.0%; Pred. No. 8.5e-253;
Matches 492; Conservative 14; Mismatches 22; Indels 1; Gaps 1;
QY 1 MSKMTSALLLIQUSCYFSSGCGKVLWPTFSHMNNITIIDELVORGEVTLASSA 60
Db 1 MSKMTSALLLIQUSCYFSSGCGKVLWPTFSHMNNITIIDELVORGEVTLASSA 60
QY 61 SISFDPSSELTLEFVPSVLTTEPEDITKOLVKRAELPKOTFMSYFSQVGEIMTFN 120
Db 61 SISFDPSSELTLEFVPSVLTTEPEDITKOLVKRAELPKOTFMSYFSQVGEIMTFN 120
QY 121 DIRKFKDIVSKKMKLQESRFDVLADAVFPFCELLAEILKIPFYVSLRFSPEYAI 180

```

Db 121 DILRFKCDIYNNKLMKKLOESRFDVVLADAVPFGEILAEILKI PFVYSIRFSPGYAI 180
Qy 181 EKHSGGLFPSPSYVVMVNSLSDQMTFIERVKNMIVYL YFEFWFOIFPMKKMDQFYSVL 240
Db 181 EKHSGGLFPSPSYVVMVNSLSDQMTFIERVKNMIVYL YFEFWFOIFPMKKMDQFYSVL 240
Qy 241 GRPTLSETMAKADIWLIRNWDFOFPHPLPNVEFGVGLHCKRAKPLPKMEEFVQSSG 300
Db 241 GRPTLSETMAKADIWLIRNWDFOFPHPLPNVEFGVGLHCKRAKPLPKMEEFVQSSG 300
Qy 301 ENGVSFSLGSMVNSTSEERANVIASALAKIPQKVLRFPGNKPDTGLNTRLYKMIPON 360
Db 301 ENGVSFSLGSMVNSTSEERANVIASALAKIPQKVLRFPGNKPDTGLNTRLYKMIPON 360
Qy 361 DLGHPKTKAFITTHGANGIYEALYHGI PMVGVPLFADQPDNIAMKAKGAAVEINFETM 420
Db 361 DLGHPKTKAFITTHGANGIYEALYHGI PMVGVPLFADQPDNIAMKAKGAAVEINFETM 420
Qy 421 TSEDLRALRVTITDSSYKENAMRLSRHHQPVKPLDRAVFWIEFWNRHKGAKHLRSAA 480
Db 421 TSEDLRALRVTITDSSYKENAMRLSRHHQPVKPLDRAVFWIEFWNRHKGAKHLRSAA 480
Qy 481 HDLTFQYHSIDVIGFLLTCVATAIFLTCKCFLSCQKFNKTRKIERE 529
Db 481 HDLTFQYHSIDVIGFLLTCVATAIFLTCKCFLSCQKFNKTRKIERE 529

```

RESULT 2
US-09-981-353-194

```

; Sequence 194, Application US/09981353
; Patent No. US20020160382A1
; GENERAL INFORMATION:
; APPLICANT: Lasek, Amy W.
; TITLE OF INVENTION: GENES EXPRESSED IN COLON CANCER
; FILE REFERENCE: PA-0038 US
; CURRENT APPLICATION NUMBER: US/09/981.353
; CURRENT FILING DATE: 2001-10-11
; NUMBER OF SEQ ID NOS: 194
; SOFTWARE: PERL Program
; SEQ ID NO 194
; LENGTH: 529
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. US20020160382A1 088078CD1
US-09-981-353-194

```

```

Query Match 83.3%; Score 2330; DB 10; Length 529;
Best Local Similarity 81.1%; Pred. No. 2.3e-225;
Matches 429; Conservative 42; Mismatches 58; Indels 0; Gaps 0;

Qy 1 MSMKWTSAALLIIQUSCYFSSGCGKVLVWPTFSHMNIKITILDELVOGHEVTVLASSA 60
Db 1 MSMKWTSAALLIIQUSCYFSSGCGKVLVWPTFSHMNIKITILDELVOGHEVTVLASSA 60
Qy 61 SIFDPPNSPSTLKEVYVSLTKTEFEDIKQVLRMAELPKDFTWMSYFSQVOEIMSTFN 120
Db 61 SIFDPPNSPSTLKEVYVSLTKTEFEDIKQVLRMAELPKDFTWMSYFSQVOEIMSTFN 120
Qy 121 DILRFKCDIYNNKLMKKLOESRFDVVLADAVPFGEILAEILKI PFVYSIRFSPGYAI 180
Db 121 DILRFKCDIYNNKLMKKLOESRFDVVLADAVPFGEILAEILKI PFVYSIRFSPGYAI 180
Qy 181 EKHSGGLFPSPSYVVMVNSLSDQMTFIERVKNMIVYL YFEFWFOIFPMKKMDQFYSVL 240
Db 181 EKHSGGLFPSPSYVVMVNSLSDQMTFIERVKNMIVYL YFEFWFOIFPMKKMDQFYSVL 240
Qy 241 GRPTLSETMAKADIWLIRNWDFOFPHPLPNVEFGVGLHCKRAKPLPKMEEFVQSSG 300
Db 241 GRPTLSETMAKADIWLIRNWDFOFPHPLPNVEFGVGLHCKRAKPLPKMEEFVQSSG 300

```

```

Qy 301 ENGVSFSLGSMVNSTSEERANVIASALAKIPQKVLRFPGNKPDTGLNTRLYKMIPON 360
Db 301 ENGVSFSLGSMVNSTSEERANVIASALAKIPQKVLRFPGNKPDTGLNTRLYKMIPON 360
Qy 361 DLGHPKTKAFITTHGANGIYEALYHGI PMVGVPLFADQPDNIAMKAKGAAVEINFETM 420
Db 361 DLGHPKTKAFITTHGANGIYEALYHGI PMVGVPLFADQPDNIAMKAKGAAVEINFETM 420
Qy 421 TSEDLRALRVTITDSSYKENAMRLSRHHQPVKPLDRAVFWIEFWNRHKGAKHLRSAA 480
Db 421 TSEDLRALRVTITDSSYKENAMRLSRHHQPVKPLDRAVFWIEFWNRHKGAKHLRSAA 480
Qy 481 HDLTFQYHSIDVIGFLLTCVATAIFLTCKCFLSCQKFNKTRKIERE 529
Db 481 HDLTFQYHSIDVIGFLLTCVATAIFLTCKCFLSCQKFNKTRKIERE 529

```

RESULT 3
US-10-057-834A-2

```

; Sequence 2, Application US/10057834A
; Publication No. US20030099960A1
; GENERAL INFORMATION:
; APPLICANT: KATLIN, MARK J.
; APPLICANT: INNOCENTI, FEDERICO
; APPLICANT: DAS, SOMA
; APPLICANT: IYER, LALITHA
; APPLICANT: SAWYER, MICHAEL
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR OPTIMIZING UGT2B7 SUBSTRATE TOXICITY
; FILE REFERENCE: ARCD:35805
; CURRENT APPLICATION NUMBER: US/10/057.834A
; CURRENT FILING DATE: 2002-08-22
; PRIOR APPLICATION NUMBER: UNKNOWN
; PRIOR FILING DATE: 2002-01-25
; NUMBER OF SEQ ID NOS: 78
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 529
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-057-834A-2

```

```

Query Match 83.3%; Score 2330; DB 15; Length 529;
Best Local Similarity 81.1%; Pred. No. 2.3e-225;
Matches 429; Conservative 42; Mismatches 58; Indels 0; Gaps 0;

Qy 1 MSMKWTSAALLIIQUSCYFSSGCGKVLVWPTFSHMNIKITILDELVOGHEVTVLASSA 60
Db 1 MSMKWTSAALLIIQUSCYFSSGCGKVLVWPTFSHMNIKITILDELVOGHEVTVLASSA 60
Qy 61 SIFDPPNSPSTLKEVYVSLTKTEFEDIKQVLRMAELPKDFTWMSYFSQVOEIMSTFN 120
Db 61 SIFDPPNSPSTLKEVYVSLTKTEFEDIKQVLRMAELPKDFTWMSYFSQVOEIMSTFN 120
Qy 121 DILRFKCDIYNNKLMKKLOESRFDVVLADAVPFGEILAEILKI PFVYSIRFSPGYAI 180
Db 121 DILRFKCDIYNNKLMKKLOESRFDVVLADAVPFGEILAEILKI PFVYSIRFSPGYAI 180
Qy 181 EKHSGGLFPSPSYVVMVNSLSDQMTFIERVKNMIVYL YFEFWFOIFPMKKMDQFYSVL 240
Db 181 EKHSGGLFPSPSYVVMVNSLSDQMTFIERVKNMIVYL YFEFWFOIFPMKKMDQFYSVL 240
Qy 241 GRPTLSETMAKADIWLIRNWDFOFPHPLPNVEFGVGLHCKRAKPLPKMEEFVQSSG 300
Db 241 GRPTLSETMAKADIWLIRNWDFOFPHPLPNVEFGVGLHCKRAKPLPKMEEFVQSSG 300
Qy 301 ENGVSFSLGSMVNSTSEERANVIASALAKIPQKVLRFPGNKPDTGLNTRLYKMIPON 360
Db 301 ENGVSFSLGSMVNSTSEERANVIASALAKIPQKVLRFPGNKPDTGLNTRLYKMIPON 360
Qy 361 DLGHPKTKAFITTHGANGIYEALYHGI PMVGVPLFADQPDNIAMKAKGAAVEINFETM 420
Db 361 DLGHPKTKAFITTHGANGIYEALYHGI PMVGVPLFADQPDNIAMKAKGAAVEINFETM 420

```

Db 361 DLGHPKTRAFITGANGIYEALYHGI PMVGIPDLFADQPDNIAMKARGAVERDNTM 420
Qy 421 TSEDLRLARLRTVITDSSYKENAMRLSRHHDPKPLDRAVFWIEFVNRHKGAKHLSSAA 480
Db 421 SSTDLNALAKRVINDPSYKENAMKLSRIQHDPKPLDRAVFWIEFVNRHKGAKHLSSAA 480
Qy 481 HDLTFWQHSIDVIGFLITCVATAIFLTKCFLESCQKFNKTRKIEKRE 529
Db 481 HDLTFWQHSIDVIGFLITCVATAIFLTKCFLESCQKFNKTRKIEKRE 529

RESULT 4
US-10-205-522-40
; Sequence 40, Application US/10205522
; Publication No. US20030077629A1
; GENERAL INFORMATION:
; APPLICANT: Penny, Laura
; APPLICANT: Galvin, Margaret
; APPLICANT: Miller, Andrew
; APPLICANT: Reidy, Michael
; TITLE OF INVENTION: Genotyping Human
; TITLE OF INVENTION: UDP-Glucuronosyltransferase 2B4 (UGT2B4), 2B7 (UGT2B7) and
; TITLE OF INVENTION: 2B15 (UGT2B15) Genes
; FILE REFERENCE: SEQ-22PRV2
; CURRENT APPLICATION NUMBER: US/10/205,522
; CURRENT FILING DATE: 2002-07-24
; PRIOR APPLICATION NUMBER: US/09/356,806
; PRIOR FILING DATE: 1999-07-20
; NUMBER OF SEQ ID NOS: 164
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 40
; LENGTH: 524
; TYPE: PRF
; ORGANISM: H. sapiens
US-10-205-522-40

Query Match 83.0%; Score 2322; DB 15; Length 524;
Best Local Similarity 81.5%; Pred. No. 1.4e-224;
Matches 427; Conservative 42; Mismatches 55; Indels 0; Gaps 0;

Qy 1 MSMKWTSAALLIOLSCYFSSGCGKVLWPTFESHMNNITILDELVORGHETVLAASA 60
Db 1 MSVMTSVILLIOLSCYFSSGCGKVLWMAEYSHMNNITILDELVORGHETVLAASA 60
Qy 61 SISDRPSPSTLKEVYVPSLTKEPEDIIKOLYKRAALEPKDTFMSYFSQVOEIMMTFN 120
Db 61 SILDPNNSSALKIEIYPTSLTKELENFIMQIKRWSDLPKDTFMSYFSQVOEIMSI 120
Qy 121 DILKPKCDIVSNKKMLKLOESRFDVVLADAVFPFGLLAELKIPVVSLSRESPGYAI 180
Db 121 DITKPKCDIVSNKKMKKQVESRFDVIFADAIFFCEBLLELNFNIPVVSLSFSGYTF 180
Qy 181 EKHSGLLFPSPSYVPMVMSLSDQMTFIERVKMIIYLYFEFWFOIFDMKKMDQFYSEVL 240
Db 181 EKHSGLLFPSPSYVPMVMSLSDQMTFIERVKMIIYLYFEFWFOIFDMKKMDQFYSEVL 240
Qy 241 GRPTLSEYMAKAIWILIRNWDQFPHPLPNVEFVGLHCKRAKPLPKMEEFVQSSG 300
Db 241 GRPTLSEYMAKAIWILIRNWDQFPHPLPNVEFVGLHCKRAKPLPKMEEFVQSSG 300
Qy 301 ENGUVVSLSGMVSNTSEERANVTASALAKIPQKVLMPFGNKPDTGLNTRLYKWI 360
Db 301 ENGUVVSLSGMVSNTSEERANVTASALAKIPQKVLMPFGNKPDTGLNTRLYKWI 360
Qy 361 DLGHPKTRAFITGANGIYEALYHGVPMVGPIFGDQLDNIAMKARGAVERDNTM 420
Db 361 DLGHPKTRAFITGANGIYEALYHGVPMVGPIFGDQLDNIAMKARGAVERDNTM 420
Qy 421 TSEDLRLARLRTVITDSSYKENAMRLSRHHDPKPLDRAVFWIEFVNRHKGAKHLSSAA 480
Db 421 SSTDLNALAKRVINDPSYKENAMKLSRIQHDPKPLDRAVFWIEFVNRHKGAKHLSSAA 480
Qy 481 HDLTFWQHSIDVIGFLITCVATAIFLTKCFLESCQKFNKTRK 524

Db 481 HDLTFWQHSIDVIGFLITCVATAIFLTKCFLESCQKFNKTRK 524

RESULT 5
US-10-205-522-113
; Sequence 113, Application US/10205522
; Publication No. US20030077629A1
; GENERAL INFORMATION:
; APPLICANT: Penny, Laura
; APPLICANT: Galvin, Margaret
; APPLICANT: Miller, Andrew
; APPLICANT: Reidy, Michael
; TITLE OF INVENTION: Genotyping Human
; TITLE OF INVENTION: UDP-Glucuronosyltransferase 2B4 (UGT2B4), 2B7 (UGT2B7) and
; TITLE OF INVENTION: 2B15 (UGT2B15) Genes
; FILE REFERENCE: SEQ-22PRV2
; CURRENT APPLICATION NUMBER: US/10/205,522
; CURRENT FILING DATE: 2002-07-24
; PRIOR APPLICATION NUMBER: US/09/356,806
; PRIOR FILING DATE: 1999-07-20
; NUMBER OF SEQ ID NOS: 164
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 113
; LENGTH: 530
; TYPE: PRF
; ORGANISM: H. sapiens
US-10-205-522-113

Query Match 77.1%; Score 2158.5; DB 15; Length 530;
Best Local Similarity 74.5%; Pred. No. 4.2e-208;
Matches 395; Conservative 60; Mismatches 74; Indels 1; Gaps 1;

Qy 1 MSMKWTSAALLIOLSCYFSSGCGKVLWPTFESHMNNITILDELVORGHETVLAASA 60
Db 1 MSKMTSVILLIOLSCYFSSGCGKVLWPTFESHMNNITILDELVORGHETVLAASA 60
Qy 61 SISDRPSPSTLKEVYVPSLTKEPEDIIKOLYKRAALEPKDTFMSYFSQVOEIMMTFN 119
Db 61 SILDPNNSSALKIEIYPTSLTKELENFIMQIKRWSDLPKDTFMSYFSQVOEIMSI 120
Qy 121 DILKPKCDIVSNKKMLKLOESRFDVVLADAVFPFGLLAELKIPVVSLSRESPGYAI 179
Db 121 DITKPKCDIVSNKKMKKQVESRFDVIFADAIFFCEBLLELNFNIPVVSLSFSGYTF 180
Qy 180 EKHSGLLFPSPSYVPMVMSLSDQMTFIERVKMIIYLYFEFWFOIFDMKKMDQFYSEV 239
Db 180 EKHSGLLFPSPSYVPMVMSLSDQMTFIERVKMIIYLYFEFWFOIFDMKKMDQFYSEV 240
Qy 240 LGRPTLSEYMAKAIWILIRNWDQFPHPLPNVEFVGLHCKRAKPLPKMEEFVQSS 299
Db 240 LGRPTLSEYMAKAIWILIRNWDQFPHPLPNVEFVGLHCKRAKPLPKMEEFVQSS 300
Qy 300 ENGUVVSLSGMVSNTSEERANVTASALAKIPQKVLMPFGNKPDTGLNTRLYKWI 359
Db 300 ENGUVVSLSGMVSNTSEERANVTASALAKIPQKVLMPFGNKPDTGLNTRLYKWI 360
Qy 360 NDILGHPKTRAFITGANGIYEALYHGVPMVGPIFGDQLDNIAMKARGAVERDNTM 419
Db 360 NDILGHPKTRAFITGANGIYEALYHGVPMVGPIFGDQLDNIAMKARGAVERDNTM 420
Qy 420 MTESEDLRLARLRTVITDSSYKENAMRLSRHHDPKPLDRAVFWIEFVNRHKGAKHLSSA 479
Db 420 MTESEDLRLARLRTVITDSSYKENAMRLSRHHDPKPLDRAVFWIEFVNRHKGAKHLSSA 480
Qy 480 HDLTFWQHSIDVIGFLITCVATAIFLTKCFLESCQKFNKTRKIEKRE 529
Db 480 HDLTFWQHSIDVIGFLITCVATAIFLTKCFLESCQKFNKTRKIEKRE 529

Patent No. US20020160382A1
GENERAL INFORMATION:
APPLICANT: Laeek, Amy W.
APPLICANT: Jones, David A.
TITLE OF INVENTION: GENES EXPRESSED IN COLON CANCER
FILE REFERENCE: PA-0038 US
CURRENT APPLICATION NUMBER: US/09/981,353
CURRENT FILING DATE: 2001-10-11
NUMBER OF SEQ ID NOS: 194
SOFTWARE: PERL Program
SEQ ID NO 166
LENGTH: 527
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc feature
OTHER INFORMATION: Incyte ID No. US20020160382A1 2434655CD1
US-09-981-353-166

Query Match 68.8%; Score 1924.5; DB 10; Length 527;
Best Local Similarity 69.9%; Pred. No. 1.5e-184;
Matches 369; Conservative 47; Mismatches 109; Indels 3; Gaps 2;

QY 2 SMKTSALLLQLSCYFSSGCGKLVLPTEFSSHMNIKITLDELVORGEVTVLASSAS 61
DB 3 SDKSALVFLQLFC-VGCGFCGKLVLPDCMSHMLNVKVLLEBLIYRGHEVTVLTHSKP 61
QY 62 ISPDNSSTLKEFVYVSLTTEFEDIIKQLVKMAELPKDTFMSYFSQVQELMTFND 121
DB 62 SLIDRKPSALKFEVVMHPDORTENEI FVDLALN--VLPGISTQSVIKLNDFFVEIRG 119
QY 122 ILRKCKDIVSNKKLMKKLOESRFDVVLADAVFPGELLAELKIPVYSLRFSGYAIE 181
DB 120 TLKMMCESFTYNQTLMKKLOETNYDVMLIDPVICGDLMAELAVFVLTLRISVGME 179
QY 182 KHSGLLFPSPSYVPMVSELSQDQMTFIERVKMIVLYFEFVFOIFDMKKADQYSEVLG 241
DB 180 RSCGKLPAPLSYVPMVGTLDRTMTFLERVNSMLSVLFHFMIQDYVHFMEEFYSKLG 239
QY 242 RPTTSETMAKADIVLRNWDQFPHPLLPNVEFVGLHCKPAKPLPKEMEFPVQSGE 301
DB 240 RPTTLCETVGAELIWLIRTYWDFEPPOYPNFEFVGLHCKPAKPLPKEMENFVQSGE 299
QY 302 NGVVFSLGSMVSNSEERANVIAALAKIPQKVLMPFDGKPDPLGLNTRLYKMIPOND 361
DB 300 DGIYVFSLSLFOVNTTEKANIISALAQIPQKVLMPYKGGKPSLTGANTRLYDNIPOND 359
QY 362 LLGHPKTAFTTHGAMNGIYEAIVHGVPMGVPIFGDQLDNIAMKAGAAVEINFKTMT 421
DB 360 LLGHPKTAFTTHGAMNGIYEAIVHGVPMGVPIFGDQLDNIAMKAGAAVEINFKTMT 419
QY 422 SEDLLRALRTVITDSSYKENAMRLSRRIHDDQVPLDRAVFIIEFVMMHKKAKHLRSAH 481
DB 420 SEDLLRALRTVITDSSYKENAMRLSRRIHDDQVPLDRAVFIIEFVMMHKKAKHLRSAH 479
QY 482 DLTWFOHYSIDVIGFLTCVATAIFLFTKCFLFCQKFNKTRKIEKE 529
DB 480 DLTWFOHYSIDVIGFLTCVATAIFLFTKCFLFCQKFNKTRKIEKE 527

RESULT 7
US-10-199-672-522
Sequence 522, Application US/10199672
Publication No. US2003014842A1

GENERAL INFORMATION:
APPLICANT: Baker, Kevin P.
APPLICANT: Chen, Jian
APPLICANT: Desnoyers, Luc
APPLICANT: Goddard, Audrey
APPLICANT: Godowski, Paul J.
APPLICANT: Gurney, Austin L.
APPLICANT: Pan, James
APPLICANT: Smith, Victoria

APPLICANT: Watanabe, Colin K.
APPLICANT: Wood, William I.
APPLICANT: Zhang, Zemin
TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
FILE REFERENCE: P3430R1C1
CURRENT APPLICATION NUMBER: US/10/199,672
CURRENT FILING DATE: 2002-07-18
PRIOR APPLICATION NUMBER: US/10/052,586
PRIOR FILING DATE: 2002-01-15
PRIOR APPLICATION NUMBER: 60/059263
PRIOR FILING DATE: 1997-09-18
PRIOR APPLICATION NUMBER: 60/059266
PRIOR FILING DATE: 1997-09-18
PRIOR APPLICATION NUMBER: 60/062250
PRIOR FILING DATE: 1997-10-17
PRIOR APPLICATION NUMBER: 60/063120
PRIOR FILING DATE: 1997-10-24
PRIOR APPLICATION NUMBER: 60/063121
PRIOR FILING DATE: 1997-10-24
PRIOR APPLICATION NUMBER: 60/063486
PRIOR FILING DATE: 1997-10-21
PRIOR APPLICATION NUMBER: 60/063540
PRIOR FILING DATE: 1997-10-28
PRIOR APPLICATION NUMBER: 60/063541
PRIOR FILING DATE: 1997-10-28
PRIOR APPLICATION NUMBER: 60/063544
PRIOR FILING DATE: 1997-10-28
Remaining Prior Application data removed - See File Wrapper or PALM.
NUMBER OF SEQ ID NOS: 612
SEQ ID NO 522
LENGTH: 527
TYPE: PRT
ORGANISM: Homo Sapien
US-10-199-672-522

Query Match 68.8%; Score 1924.5; DB 12; Length 527;
Best Local Similarity 69.9%; Pred. No. 1.5e-184;
Matches 369; Conservative 47; Mismatches 109; Indels 3; Gaps 2;

QY 2 SMKTSALLLQLSCYFSSGCGKLVLPTEFSSHMNIKITLDELVORGEVTVLASSAS 61
DB 3 SDKSALVFLQLFC-VGCGFCGKLVLPDCMSHMLNVKVLLEBLIYRGHEVTVLTHSKP 61
QY 62 ISPDNSSTLKEFVYVSLTTEFEDIIKQLVKMAELPKDTFMSYFSQVQELMTFND 121
DB 62 SLIDRKPSALKFEVVMHPDORTENEI FVDLALN--VLPGISTQSVIKLNDFFVEIRG 119
QY 122 ILRKCKDIVSNKKLMKKLOESRFDVVLADAVFPGELLAELKIPVYSLRFSGYAIE 181
DB 120 TLKMMCESFTYNQTLMKKLOETNYDVMLIDPVICGDLMAELAVFVLTLRISVGME 179
QY 182 KHSGLLFPSPSYVPMVSELSQDQMTFIERVKMIVLYFEFVFOIFDMKKADQYSEVLG 241
DB 180 RSCGKLPAPLSYVPMVGTLDRTMTFLERVNSMLSVLFHFMIQDYVHFMEEFYSKLG 239
QY 242 RPTTSETMAKADIVLRNWDQFPHPLLPNVEFVGLHCKPAKPLPKEMEFPVQSGE 301
DB 240 RPTTLCETVGAELIWLIRTYWDFEPPOYPNFEFVGLHCKPAKPLPKEMENFVQSGE 299
QY 302 NGVVFSLGSMVSNSEERANVIAALAKIPQKVLMPFDGKPDPLGLNTRLYKMIPOND 361
DB 300 DGIYVFSLSLFOVNTTEKANIISALAQIPQKVLMPYKGGKPSLTGANTRLYDNIPOND 359
QY 362 LLGHPKTAFTTHGAMNGIYEAIVHGVPMGVPIFGDQLDNIAMKAGAAVEINFKTMT 421
DB 360 LLGHPKTAFTTHGAMNGIYEAIVHGVPMGVPIFGDQLDNIAMKAGAAVEINFKTMT 419
QY 422 SEDLLRALRTVITDSSYKENAMRLSRRIHDDQVPLDRAVFIIEFVMMHKKAKHLRSAH 481
DB 420 SEDLLRALRTVITDSSYKENAMRLSRRIHDDQVPLDRAVFIIEFVMMHKKAKHLRSAH 479
QY 482 DLTWFOHYSIDVIGFLTCVATAIFLFTKCFLFCQKFNKTRKIEKE 529

Db 480 DLTWFQHSIDVIGFLLTCAATAFLFTKCFLSCQKFNKTRKIEKRE 527

RESULT 8
US-10-187-749-522

```

: Sequence 522, Application US/10187749
: Publication No. US20030153036A1
: GENERAL INFORMATION:
: APPLICANT: Baker, Kevin P.
: APPLICANT: Chen, Jian
: APPLICANT: Desnoyers, Luc
: APPLICANT: Goddard, Audrey
: APPLICANT: Godowski, Paul J.
: APPLICANT: Gurney, Austin L.
: APPLICANT: Pan, James
: APPLICANT: Smith, Victoria
: APPLICANT: Watanabe, Colin K.
: APPLICANT: Wood, William I.
: APPLICANT: Zhang, Zhenli
: TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
: FILE REFERENCE: P3430R1C1
: CURRENT APPLICATION NUMBER: US/10/187,749
: PRIORITY FILING DATE: 2002-07-01
: PRIOR APPLICATION NUMBER: US/10/052,586
: PRIOR FILING DATE: 2002-01-15
: PRIOR APPLICATION NUMBER: 60/059263
: PRIOR FILING DATE: 1997-09-18
: PRIOR APPLICATION NUMBER: 60/059266
: PRIOR FILING DATE: 1997-09-18
: PRIOR APPLICATION NUMBER: 60/062250
: PRIOR FILING DATE: 1997-10-17
: PRIOR APPLICATION NUMBER: 60/063120
: PRIOR FILING DATE: 1997-10-24
: PRIOR APPLICATION NUMBER: 60/063121
: PRIOR FILING DATE: 1997-10-24
: PRIOR APPLICATION NUMBER: 60/063486
: PRIOR FILING DATE: 1997-10-21
: PRIOR APPLICATION NUMBER: 60/063540
: PRIOR FILING DATE: 1997-10-28
: PRIOR APPLICATION NUMBER: 60/063541
: PRIOR FILING DATE: 1997-10-28
: PRIOR APPLICATION NUMBER: 60/063544
: PRIOR FILING DATE: 1997-10-28
: Remaining seq Application data removed - See File Wrapper or PAM.
: NUMBER OF SEQ ID NOS: 612
: SEQ ID NO 522
: LENGTH: 527
: TYPE: PRT
: ORGANISM: Homo Sapien
: US-10-187-749-522

```

| | | | | |
|-----------------------|-----------------|--------------------|----------|------------|
| Query Match | 68.8% | Score 1924.5 | DB 12 | Length 527 |
| Best Local Similarity | 69.9% | Pred. No. 1.5e-184 | | |
| Matches 369 | Conservative 47 | Mismatches 109 | Indels 3 | Gaps 2 |

| | | | |
|----|-----|--|-----|
| QY | 2 | SMKMTALHIIQJSCYFSSGCGKVLVMPTEESHNNITITIDELVQGHETVLASSAS | 61 |
| Db | 3 | SBSKALVFLIIQJFC-VGGCGFGKVLVMPDCDSHMLNKKVILIEELVIGHEVYVLTSHKP | 61 |
| QY | 62 | ISFDPSPTLTKEEYVPSVLTKEFEDIKOLVKMAELPKDTFWSYFSQVOELMTMFND | 121 |
| Db | 62 | SLIDYRKPSALKEVEVHMPQDRTEENELFVDLALN-VLPGLSTQSVIYKLNDFVEIRG | 119 |
| QY | 122 | ILAKFCDIYSNKKMLMKLOESRFVDVADAVFPFGELLAELIKLPPYYSIRFSGVAIE | 181 |
| Db | 120 | TLKMWCESPYNOTLTKMKLOETVYDPMALIDPVIKPGDMLAELIADVPVLTLRISVGME | 179 |
| QY | 182 | KHSGLGLFPSPYVPVWSELSDDMTIEEVKMIYIVLFEFPOFEDMKKXDFSEVLTG | 241 |
| Db | 180 | RSCGKLPAPLSYVPVMPGTGLTDMPTLEEVKSMISLVFHPITQUDYDHWEEFSAKLG | 239 |

| | | | |
|----|-----|---|----------|
| QY | 242 | PPTLSEMAKADILINRYMDFQEPHPLPLVPEFGLCHCKPAKLPKMEMEFVSSGE | 301 |
| Db | 240 | RPTLCETVGAKEIMLIRTYMDFEFPORQPNPEFVGGLHCKPAKLPKMEMEFVSSGE | 299 |
| QY | 302 | NGVVVFSLGSWVSNTSEBRANVIASALAKIPOKVLWRFDGKPKDTGLNTRLYKWI | POND 361 |
| Db | 300 | DGIIVFVSGISLFQONTTEEKANIIASALAQIPOKVLWRYKGGKPKSTLGANTRLTDWI | POND 359 |
| QY | 362 | LLGHKTAFLTHGGMNGIYEAIYHGVPMVGPVIFGDOLDNIAHMAKGAAYENFKTMT | 421 |
| Db | 360 | LLGHKTAFLTHGGMNGIYEAIYHGVPMVGPVIFGDOLDNIAHMAKGAAYENFKTMT | 419 |
| QY | 422 | SEDLRLALRTVITDSSYKENAMRLSRIHHDOVPKPLDRAPVWIEFVRRHGAKGLRSSAH | 481 |
| Db | 420 | SEDLRLALRTVITDSSYKENAMRLSRIHHDOVPKPLDRAPVWIEFVRRHGAKGLRSSAH | 479 |
| QY | 482 | DLTWFOHNSIDVIGFLTCVATAIPLFKCFLFSQCKFKTKTIERE | 529 |
| Db | 480 | DLTWFOHNSIDVIGFLTCVATAIPLFKCFLFSQCKFKTKTIERE | 527 |

RESULT 9
US-10-194-457-522

```

? Sequence 522, Application US/101994457
? Publication No. US20030153037A1
? GENERAL INFORMATION:
? APPLICANT: Baker, Kevin P.
? APPLICANT: Chen, Jian
? APPLICANT: Desnoyers, Luc
? APPLICANT: Goddard, Audrey
? APPLICANT: Godowski, Paul J.
? APPLICANT: Gurney, Austin L.
? APPLICANT: Pan, James
? APPLICANT: Smith, Victoria
? APPLICANT: Watanabe, Colin K.
? APPLICANT: Wood, William I.
? APPLICANT: Zhang, Zemin
? TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
? TITLE OF INVENTION: ACIDS ENCODING THE SAME
? FILE REFERENCE: P3430R1C296
? CURRENT APPLICATION NUMBER: US/10/194,457
? CURRENT FILING DATE: 2002-07-11
? PRIOR APPLICATION NUMBER: 10/052586
? PRIOR FILING DATE: 2002-01-15
? PRIOR APPLICATION NUMBER: 60/059263
? PRIOR FILING DATE: 1997-09-18
? PRIOR APPLICATION NUMBER: 60/059266
? PRIOR FILING DATE: 1997-09-18
? PRIOR APPLICATION NUMBER: 60/062250
? PRIOR FILING DATE: 1997-10-17
? PRIOR APPLICATION NUMBER: 60/063120
? PRIOR FILING DATE: 1997-10-24
? PRIOR APPLICATION NUMBER: 60/063121
? PRIOR FILING DATE: 1997-10-24
? PRIOR APPLICATION NUMBER: 60/063486
? PRIOR FILING DATE: 1997-10-21
? PRIOR APPLICATION NUMBER: 60/063540
? PRIOR FILING DATE: 1997-10-28
? PRIOR APPLICATION NUMBER: 60/063541
? PRIOR FILING DATE: 1997-10-28
? PRIOR APPLICATION NUMBER: 60/063544
? PRIOR FILING DATE: 1997-10-28
? NUMBER OF SEQ ID NOS: 612
? SEQ ID NO 522
? LENGTH: 527
? TYPE: PRT
? ORGANISM: Homo Sapien
? US-10-194-457-522

```

| | | | | |
|-----------------------|-----------------|--------------------|----------|------------|
| Query Match | 68.8% | Score 1924.5 | DB 12 | Length 527 |
| Best Local Similarity | 69.9% | Pred. NO. 1.5e-184 | | |
| Matches 369 | Conservative 47 | Mismatches 109 | Indels 3 | Gaps 2 |

```

Qy 2 SMKQTSALLIQLQSCYFSSGSCGKLVMPTEFSHMNITIIIDELVQGHVYLVASSAS 61
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 3 SDKALVYELLQFLC-VGCGFCGKYLVMPCDMSHMLNVAKILBELLVRHGEVYLVHNSRP 61
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

Qy 62 ISPDNSPSTLKFEVYPVSLTTEBEDIIKQLVKMAELPKOTFMSYFSGVQGEIMTMD 121
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 62 SLIDRKKSALLFEVVMHPQDRTTEBNEIRVDLALN--VLPGSLTMSQSVIKLNDPVEIRG 119
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

Qy 122 ILRKXCQIVSKKMLMKKQJQESRPVDLDAVFPFGELLAELKIPFVSLRSPGAYLE 181
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 120 TLKMMCESFIYQTLMKLOJENNYDMLIDPVIPOCDLMAELLAVFLVTLIRSVGGNNE 179
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

Qy 182 KHSGLLFPSPSYVPMVMSLSQMTFIEBKNNIYVLVEPWFQITDKKMDQFYSEVLG 241
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 180 RSCGKLPLPLSVYPVPMITGLDTRMTFLERVKMSMLSVLPHFWIQQDDYHMEBEFYKALG 239
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

Qy 242 RPTTSEMAKKADIWLIRVYMDPOFPHPLPNVEFVGJLHCKRAKPLPKMEBEFYOSSGE 301
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 240 RPTTLCETVYKKEIMLIRYTMDEFPQRPQPNFEPVGGJLHCKRAKPLPKMEBEFYOSSGE 239
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

Qy 302 NGVVVFSLGSWVNTSEERANVITASALAKI POKVLVRFQGNKPDYTLGLNTRLYKMI POND 361
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 300 DGIIVFVSLGSLFQNTVTEERKANIITASALAOI POKVLVRYKGGKSPSTIGANTRLYDMI POND 359
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

Qy 362 LIGHKRTKAFITHGNGNGIYEALIVHGVPMVGPVIFBDQDNIINMKAKGAAYVEINFKTM 421
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 360 LIGHKRTKAFITHGNGNGIYEALIVHGVPMVGPVIFBDQDNIINMKAKGAAYVEINFKTM 419
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

Qy 422 SEDLLRALTVITDSSYKENAMKLSRIHDDQPKPLDRAVFWIEFWVRHKGAHLRSAH 481
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 420 SEDLLRALTVITDSSYKENAMKLSRIHDDQPKPLDRAVFWIEFWVRHKGAHLRSAH 479
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

Qy 482 DLTWEQHSIDVIGFLLTCVATPIELFTCCGFELSSCKPNTKRIKRE 529
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 480 DLTWEQHSIDVIGFLLTCVATPIELFTCCGFELSSCKPNTKRIKRE 527
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

```

```

US-10-184-642-522
RESULT 10
; Sequence 522, Application US/10184642
; Publication No. US20030157635A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; TITLE OF INVENTION: ACIDS ENCODING THE SAME
; FILE REFERENCE: P3430R1C194
; CURRENT APPLICATION NUMBER: US/10/184,642
; CURRENT FILING DATE: 2002-06-27
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 522
; LENGTH: 527
; TYPE: prt
; ORGANISM: Homo Sapien
US-10-184-642-522

```

```

Query Match 68.8%; Score 1924.5; DB 12; Length 527;
Best Local Similarity 69.9%; Pred. No. 1.5e-184;
Matches 369; Conservative 47; Mismatches 109; Indels 3; Gaps 2
QY 2 SNAKTSALLLIQISCFSSGSCGKLVLPTEFSPHMNNITLIDELVQRGHEVTYLAASSAS 61
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

```

Db      3  SDRKALVFLLLQLQFC-VGCGFCGKLVWPCDMSHMLNFKVILLEEILVGHETVLTLSHXP 61
Qy      62  ISFPDNGSPTLKFEVYVYVSLFKTEBEDIIXOLVYKRWALBKOTEMYSFOVGEIMWTFMD 121
Db      62  SLIDYRKPSAKFEVWHPQDRTENEELFVDALN--VLPGLSTQSYIKLMDFVEIENG 119
Qy      122  ILARFKCDIVANKLMLKQLESRPBVVLADAVFPPGELLAEILKIPVYSLSRSPGAYLE 181
Db      120  TLKMWCSFIIHQTLMLKKLQETNVDMVLLDPVIFCGDLMAEILAVPVLTLRISVGNNME 179
Qy      182  KHSGGLLPBSYVYVWVMSLSDQMTFIRYVKMIVLVLYFEFPOFLFDMKKMOOFSEVUG 241
Db      180  RSCGGLPRLPSIVYVPMPTGLTDRMTFLERVKNSMLSTVFHEMTIOQDYHFMEEFYSKALG 239
Qy      242  RPTLTSEMAKADIVILIRNYMDPOCFPHLLPVNEFVGGILHCKRDKAPLPRKEMEEFYQSSGE 301
Db      240  RPTLTCEIVGKAELIMLIRTYWDFEFPQYQRPFRFEVGGILHCKRPAVALPREMENFVQSSGE 299
Qy      302  NGVVVFSTGSMVSNSTSEBRANVYASALAKIPQKVLNRPDGNKRPDTLGLNTRLYKMTPOND 361
Db      300  DGIWVFSIGSIFONTTEKANIISALAQIPQKVLNRYKGGKRPSTLGATRLTYLDMTIPOND 359
Qy      362  LLGHPTKATPITTHGGMNGIYEALVYHGVPMVGPVIFGDDQDNIANHKAKAAVAEINFKTMT 421
Db      360  LLGHPTKATPITTHGGMNGIYEALVYHGVPMVGPVIFGDDQDNIANHKAKAAVAEINFKTMT 419
Qy      422  SEDLLRALRTVITDSSYKENAMRLSRIHHDQVPLDRAVPMIEFVYMRHKGAKHLSAAH 481
Db      420  SEDLLRALRTVITDSSYKENAMRLSRIHHDQVPLDRAVPMIEFVYMRHKGAGHLSAAH 479
Qy      482  DLTWFOHNSIDVIGLTCVATAITLFTKCFCLFSCOKFKTKIEKRE 529
Db      480  DLTWFOHNSIDVIGLTCVATAITLFTKCFCLFSCOKFKTKIEKRE 527

```

```

RESULT 11
US-10-196-747-522
; Sequence 522, Application US/10196747
; Publication No. US20030162250A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zhenli
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; ACIDS ENCODING THE SAME
; FILE REFERENCE: P3430R1C346
; CURRENT APPLICATION NUMBER: US/10/196,747
; CURRENT FILING DATE: 2002-07-16
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 522
; LENGTH: 527
; TYPE: prt
; ORGANISM: Homo Sapien
US-10-196-747-522

```

| | | | | |
|-----------------------|-------|--------------------|-------|----------------------------------|
| Query Match | 68.8% | Score 1224.5 | DB 12 | Length 527 |
| Best Local Similarity | 69.9% | Pred. No. 1.5e-184 | | |
| Matches | 369 | Conservative | 47 | Mismatches 109; Indels 3; Gaps 2 |

| | | | |
|----|----|--|-----|
| QY | 2 | SNKMTSALLIQLSCYFSSGCGKLVNPTSESHMNIKTIIDELVQSGHEVTVLASSAS | 61 |
| | | | |
| | | | |
| Db | 3 | SKXSLVFLIDLPFC-VGCGFCGKLVNPTCDSHNLNVKYLIELLVKQHEVTVLTHSKP | 61 |
| | | | |
| QY | 62 | ISFDNPSPTLKEFVYVSLTITREEDIIKQLVKMAELPKPTFMSYFSQVQEIIMTTFND | 121 |

```
Db 62 SLIDYRKPSALKFEVVMPODRTEENEIFVDLALN--VLPGLSTWQSVIKLNDPFEVIRG 119
Qy 122 ILRKFCDDIVSNKKLMKKLOESRFDVVLADAVFPFGELLAELKIPFVSLRSPGYAIE 181
Db 120 TLKMMCESFIYNGTLMKKLOETNYDVLIDPVI PCGDLMAELLAVPVLTLRISVGNGME 179
Qy 182 KHSGLLFPSPSYVPMVMSLSDQMTPIERVKNMIVLYFEFPOIFPMKKWDQFSEVLG 241
Db 180 RSCGKLPAPIASYVPVPMPTGLTDRMTFLERVKNSMLSTLFHFMIODYHHEEYFSAIG 239
Qy 242 RPTLTSETMAKADIWLIRNWDFOFPHPLPNVEFVGLHCCKPAKLPKEMEERVOSSGE 301
Db 240 RPTLTCEYGAKEIWLIRTYWDFEPQYQNFEPVGLHCCKPAKLPKEMENVOSSGE 299
Qy 302 NGVVFSLGSMVNTSEERANVIALAKIPQKYLRFDPDKPPTLGLNTRLYKWIQND 361
Db 300 DGIIVFSLGSLFQNVTEKANIIASALAOIPQKYLWRYKKGKSTLGANTRLYDWMIPND 359
Qy 362 LIGHPKTKAFITHGMMGIYEAIYHGVPMGVPIFGQDLNIAHMKAGAAVEINFXTMT 421
Db 360 LIGHPKTKAFITHGMMGIYEAIYHGVPMGVPIFGQDLNIAHMKAGAAVEINFXTMT 419
Qy 422 SEDLLRALRTVITDSSYKENAMRLSRTHDQPVKPLDRAVFWIEFVNRHKGAKHLRSAH 481
Db 420 SEDLLRALRTVITDSSYKENAMRLSRTHDQPVKPLDRAVFWIEFVNRHKGAKHLRSAH 479
Qy 482 DLTWFOHYSIDVIGFLLTCVATAIFLTKCFLSCQKFNTRKIEKRE 529
Db 480 DLTWFOHYSIDVIGFLLTCVATAIFLTKCFLSCQKFNTRKIEKRE 527
```

RESULT 12

```
US-10-173-689-522
; Sequence 522, Application US/10173689
; Publication No. US2003016610A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3430R1C10
; CURRENT APPLICATION NUMBER: US/10/173,689
; CURRENT FILING DATE: 2002-06-17
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 522
; LENGTH: 527
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-173-689-522
```

Query Match 68.8%; Score 1924.5; DB 12; Length 527;
Best Local Similarity 69.9%; Pred. No. 1.5e-184;
Matches 369; Conservative 47; Mismatches 109; Indels 3; Gaps 2;

```
Qy 2 SMKWTSAALLLIQSCYFSSGCGKVLVMPTEFSHMNIKITILDELVORGHEVTLASSAS 61
Db 3 SDRKALVFLILQLFC-VGCGFCGKVLVMPCDMSHMLNVKILBELIYRGHEVTLASSAS 61
Qy 62 ISFDPNSPSTLKEFVYVSLTKTEFEDIIKQLVYRWAELEPDTFWSYFSQVOEIMTFFND 121
Db 62 SLIDYRKPSALKFEVVMPODRTEENEIFVDLALN--VLPGLSTWQSVIKLNDPFEVIRG 119
```

```
Qy 122 ILRKFCDDIVSNKKLMKKLOESRFDVVLADAVFPFGELLAELKIPFVSLRSPGYAIE 181
Db 120 TLKMMCESFIYNGTLMKKLOETNYDVLIDPVI PCGDLMAELLAVPVLTLRISVGNGME 179
Qy 182 KHSGLLFPSPSYVPMVMSLSDQMTPIERVKNMIVLYFEFPOIFPMKKWDQFSEVLG 241
Db 180 RSCGKLPAPIASYVPVPMPTGLTDRMTFLERVKNSMLSTLFHFMIODYHHEEYFSAIG 239
Qy 242 RPTLTSETMAKADIWLIRNWDFOFPHPLPNVEFVGLHCCKPAKLPKEMEERVOSSGE 301
Db 240 RPTLTCEYGAKEIWLIRTYWDFEPQYQNFEPVGLHCCKPAKLPKEMENVOSSGE 299
Qy 302 NGVVFSLGSMVNTSEERANVIALAKIPQKYLRFDPDKPPTLGLNTRLYKWIQND 361
Db 300 DGIIVFSLGSLFQNVTEKANIIASALAOIPQKYLWRYKKGKSTLGANTRLYDWMIPND 359
Qy 362 LIGHPKTKAFITHGMMGIYEAIYHGVPMGVPIFGQDLNIAHMKAGAAVEINFXTMT 421
Db 360 LIGHPKTKAFITHGMMGIYEAIYHGVPMGVPIFGQDLNIAHMKAGAAVEINFXTMT 419
Qy 422 SEDLLRALRTVITDSSYKENAMRLSRTHDQPVKPLDRAVFWIEFVNRHKGAKHLRSAH 481
Db 420 SEDLLRALRTVITDSSYKENAMRLSRTHDQPVKPLDRAVFWIEFVNRHKGAKHLRSAH 479
Qy 482 DLTWFOHYSIDVIGFLLTCVATAIFLTKCFLSCQKFNTRKIEKRE 529
Db 480 DLTWFOHYSIDVIGFLLTCVATAIFLTKCFLSCQKFNTRKIEKRE 527
```

RESULT 13

```
US-10-173-690-522
; Sequence 522, Application US/10173690
; Publication No. US2003016610A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3430R1C9
; CURRENT APPLICATION NUMBER: US/10/173,690
; CURRENT FILING DATE: 2002-06-17
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 522
; LENGTH: 527
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-173-690-522
```

Query Match 68.8%; Score 1924.5; DB 12; Length 527;
Best Local Similarity 69.9%; Pred. No. 1.5e-184;
Matches 369; Conservative 47; Mismatches 109; Indels 3; Gaps 2;

```
Qy 2 SMKWTSAALLLIQSCYFSSGCGKVLVMPTEFSHMNIKITILDELVORGHEVTLASSAS 61
Db 3 SDRKALVFLILQLFC-VGCGFCGKVLVMPCDMSHMLNVKILBELIYRGHEVTLASSAS 61
Qy 62 ISFDPNSPSTLKEFVYVSLTKTEFEDIIKQLVYRWAELEPDTFWSYFSQVOEIMTFFND 121
Db 62 SLIDYRKPSALKFEVVMPODRTEENEIFVDLALN--VLPGLSTWQSVIKLNDPFEVIRG 119
Qy 122 ILRKFCDDIVSNKKLMKKLOESRFDVVLADAVFPFGELLAELKIPFVSLRSPGYAIE 181
Db 120 TLKMMCESFIYNGTLMKKLOETNYDVLIDPVI PCGDLMAELLAVPVLTLRISVGNGME 179
```

```

Oy 182 KSGGGLPEPSPVPVPMSELSDMTEIERKJMIYVLYEPEFMQIPMKMDQFYSVLG 241
Oy 180 RSCGKLPAULSTVVPVMTGLTDMTELERKJMSVLFPFHWQDDYHFEWEEYSALG 239
Oy 242 RPTTSETMAKADIMWLIRNWMDEQFPHLLPNVEFVGJLHCKRCKAPLPKMEEFVSSGE 3010
Oy 240 RPTTICEIYKKEIWLIRTYMDEFEQPOPNPEFVGJLHCKRCKAKALPKMENFVSSGE 2399
Oy 302 NGVVVFSLSGWSVNTSEERANVTASALAKI POKVLWRFPDGNKDDTGLNTRLYKWI POND 3610
Oy 300 DGIIVFSGLSLQNVNTEERANVTASALAQ POKVLWRKYKCKDSTIGANTRLYDWI POND 3560
Oy 362 LIGHKPTKAFITHGKNGCIYEALYHGVPMVGVP IFGDOLDNIAMKAKGAABINFTMT 421
Oy 360 LIGHKPTKAFITHGKNGCIYEALYHGVPMVGVP IFGDOLDNIAMKAKGAABINFTMT 419
Oy 422 SEDLLRALPTVTITDSSYKENAMLSIRIHDOPKPLDRAPFWALFVWRHGHGALRSAAH 481
Oy 420 SEDLLRALPTVTITDSSYKENAMLSIRIHDOPKPLDRAPFWALFVWRHGHGALRSAAH 479
Oy 462 DLTWFOHYSIDVIGFLTCAVATAIPLPTCKFLPSCOKFNKTRIKRE 529
Oy 460 DLTWFOHYSIDVIGFLTCAVATAIPLPTCKFLPSCOKFNKTRIKRE 527

```

```

RESULT 14
US-10-173-691-522
; Sequence 522, Application US/10173691
; Publication No. US20030166106A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; TITLE OF INVENTION: ACIDS ENCODING THE SAME
; FILE REFERENCE: P3430R1C13
; CURRENT APPLICATION NUMBER: US/10/173,691
; CURRENT FILING DATE: 2002-06-17
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO: 522
; LENGTH: 527
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-173-691-522

Query Match      68.8%; Score 1924.5; DB 12; Length 527;
Beet Local Similarity 69.9%; Pred. No. 1.5e-184;
Matches 369; Conservative 47; Mismatches 109; Indels 3; Gaps 2

2  SMKMTSALLILQSCYFSSGCGKVLWPTBESHMNNITILDELVORGEVTLASSAS 61
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
3  SDKSALVFLILQLFC-VGCGFCGKVLWPCDMSHMLNVLKILIEILVRGEVTLVLSKP 61
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :

62  ISFDNSPTLKFVYVPSVLTITEPEDITKQLVKRAELPKOTFMSYFSQVOEIMTFND 121
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
62  SLIDRRKPKALFEVVMHPQDRTEENEIIVDALN--VLPGLSTWQSVIKINDFVEIRG 119
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :

122  ILRKFCKDVSNNKLTMLKQOESRPDVLADAVFPFGELIAELKTIPIVYSLRFSPGAYE 181
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
120  TLKMMCESTIVQTLMKLQLEINNYDMLIDPIVPCDDLAELIALLVAFVLTILKISVGANNE 179
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
182  KHSGLLPSPSVPVVMSLELDQMFTIEVKNNIYLVYEFNFQJLFDKMKWDQFTSEVIG 241
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

| | | | |
|----|-----|---|-----|
| Db | 180 | RSCGLPAPLBSVVPYPMGTGLDRMTFLEKVKXMSLTVLFHFIQDYDYHFMEEFYSKALG | 239 |
| QY | 242 | RPTTISETMARADIMLRINMYDFOCPHPLIPNVEYVGSIHCKPAPKLPREMEFVSSGE | 301 |
| Db | 240 | RPTTLCEIVGAEIWLIRTYWDFEPPQOPQPFEPFGGJHCKPAPKLPREMEFVSSGE | 299 |
| QY | 302 | NGVVVFSIGSNWSNTESEERANVIAALAKIPKVLMPEDGNKPDITGLNTRLYKMPOND | 361 |
| Db | 300 | DGIWVFSIGSLFQNVNTEERANIIASALAQIPKVLMPRYKGGKPSFLGANTRLDYMPIOND | 359 |
| QY | 362 | LLGHKTKAFTTHGSMNGIYEAIYHGVPMVGPPIGDOLDNIAHMKAKAAVEINRKTMT | 421 |
| Db | 360 | LLGHKTKAFTTHGSMNGIYEAIYHGVPMVGPPIGDOLDNIAHMKAKAAVEINRKTMT | 419 |
| QY | 422 | SEDLRLALRTVITDSSYKENAMRLSRIRHDDQVKKPLDRAVFWIEFVRRHGAGHLRSAAH | 481 |
| Db | 420 | SEDLRLALRTVITDSSYKENAMRLSRIRHDDQVKKPLDRAVFWIEFVRRHGAGHLRSAAH | 479 |
| QY | 482 | DLTWQHSIDVIGELTLCVAATAIPLFKPCFLPSCOKFKTKTIERE | 529 |
| Db | 480 | DLTWQHSIDVIGELTLCVAATAIPLFKPCFLPSCOKFKTKTIERGE | 527 |

```

RESULT 15
US-10-173-692-522
; Sequence 522, Application US/10173692
; Publication No. US2003016188A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Guirey, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; TITLE OF INVENTION: ACIDS ENCODING THE SAME
; FILE REFERENCE: P343081C20
; CURRENT APPLICATION NUMBER: US/10/173.692
; CURRENT FILING DATE: 2002-06-17
; Prior Application removed - See file wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 522
; LENGTH: 527
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-173-692-522

Query Match      68.8% Score 1924.5; DB 12; Length 527;
Best Local Similarity 69.9%; Pred. No. 1.5e-184;
Matches 369; Conservative 47; Mismatches 109; Indels 3; Gaps 2

      2  SMKWTALLLLQLSCYFSSGCGKLVMPTEFSSHMWNKTIIDELVORGHEVTVLASSAS 61
Db      3  SDKSLVLLQLQFRC-VCCGFCGKLVMPCDMSHMLNKVLLBELVYGHEVTVLTGHPK 61
      62  ISPDENSESTLKFEVYVPLSLTKTEFEDIIKOLVKKMAELPKDTFMSYFSQVQELMTWTD 121
      62  SLIDYRKSSALKFEVYVHNPDRTEENEIFVLDALN-VLPGISTWQSVIKINDFVEVIRG 119
      122  ILRKCKRIVSNKMLMKRLQDSRPVVLADAVFPFGCLLAEILKIPFYVLSRFSQAYLE 181
      120  TLKMCSEFFITNQTLMKRLQETNVDVMLIDPILPCGDIMAEELVAPVLLTIRISVGNNH 179
      182  KHSGLLPSPSPVPVPMVMSLSDOMTFIERVKNMIVLVLYFEFMPQIIFDKKKNDQFYSEVLG 241
      180  RSCGGLPRLPSVVPVPMGLTDRMTFFLERVKNMSLVLFHFMQIDYDVFHMFEEFYSKALG 239
      242  RPTLTSETMAKADIVLIRNWDFOFPHFLPNVEFVGGLHCKRAKRLPKENKEEFVQSSGE 301

```

```
Db      240 RPTTLCETVGKABIWLI RTYWD FEFPOYPQPNFEFVGGLHCKPAKALPKEMENFVQSSGE 299
Qy      302 NGVVVFSLSGWSVNTSEERANVITASALAKI POKYLMRFDDGK PPTLGINTRLYKTI POND 361
Db      300 DGI VFSLSGLFQVTEBKANI IASALAOI POKYLMRYKGGKPPSTLGANTRLYDIMPOND 359
Qy      362 LIGHPTKAPITTHGMMNGIYEAIYHGVPMVGP IFGDOLDNIAHMKAKGAABEINFKTM 421
Db      360 LIGHPTKAPITTHGMMNGIYEAIYHGVPMVGP IFGDOLDNIAHMKAKGAABEINFKTM 419
Qy      422 SEDLLRALRTVITDSSYKENMRLSRIHHDQPVKPLDRAVPWIEFVNRHKGAKHLRSAH 481
Db      420 SEDLLRALRTVITDSSYKENMRLSRIHHDQPVKPLDRAVPWIEFVNRHKGAKHLRSAH 479
Qy      482 DLTWFOHYSIDVIGFLTVCVATAI FLTKCFLFSCQKFNKTRKIEKRE 529
Db      480 DLTWFOHYSIDVIGFLTVCVATAI FLTKCFLFSCQKFNKTRKIEKRE 527
```

Search completed: December 5, 2003, 09:55:22
Job time : 38 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 5, 2003, 09:42:24 ; Search time 47 Seconds
(without alignments)
1786.518 Million cell updates/sec

Title: US-09-980-729B-5
Perfect score: 2798
Sequence: 1 MSMKMTSALLILQLSCYFSS.....KCFLFSCQKFNKTRIEKRE 529

Scoring table: BIOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues
Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: A_Geneseq.19Jun03.*
2: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1980.DAT.*
3: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1981.DAT.*
4: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1982.DAT.*
5: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1983.DAT.*
6: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1984.DAT.*
7: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1985.DAT.*
8: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1986.DAT.*
9: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1987.DAT.*
10: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1988.DAT.*
11: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1989.DAT.*
12: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1990.DAT.*
13: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1991.DAT.*
14: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1992.DAT.*
15: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1993.DAT.*
16: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1994.DAT.*
17: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1995.DAT.*
18: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1996.DAT.*
19: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1997.DAT.*
20: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1998.DAT.*
21: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1999.DAT.*
22: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA2000.DAT.*
23: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA2001.DAT.*
24: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA2002.DAT.*
25: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA2003.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|--------|-------------|--------|-------|-------------|
| 1 | 2798 | 100.0 | 529 | 21 | AA28677 |
| 2 | 2602.5 | 93.0 | 528 | 21 | AA78933 |
| 3 | 2584.5 | 92.4 | 540 | 22 | ABG05525 |
| 4 | 2503 | 89.5 | 529 | 23 | AA022794 |
| 5 | 2330 | 83.3 | 529 | 23 | AA28617 |
| 6 | 2325 | 83.1 | 533 | 22 | ABG05523 |
| 7 | 2322 | 83.0 | 524 | 21 | AA78934 |
| 8 | 2292 | 81.9 | 529 | 22 | AA202188 |
| 9 | 2233 | 79.8 | 532 | 22 | ABG05524 |

| | | | | | | |
|----|--------|------|-----|----|----------|---------------------|
| 10 | 2169.5 | 77.5 | 530 | 19 | AAW47126 | Uridine diphospho- |
| 11 | 2158.5 | 77.1 | 530 | 21 | AA78933 | Human UDP-glucuron |
| 12 | 2158.5 | 77.1 | 530 | 24 | ABJ19806 | Androgen-independe |
| 13 | 1924.5 | 68.8 | 527 | 22 | AAU29284 | Human PRO polypept |
| 14 | 1924.5 | 68.8 | 527 | 23 | AAU77927 | Human drug-metabol |
| 15 | 1924.5 | 68.8 | 527 | 23 | AAE15434 | Human drug-metabol |
| 16 | 1924.5 | 68.8 | 527 | 24 | AAE15434 | Human NOVX polypep |
| 17 | 1924.5 | 68.8 | 527 | 24 | ABU71372 | Human PRO6239 prot |
| 18 | 1924.5 | 68.8 | 527 | 24 | ABU65889 | Human secreted/tira |
| 19 | 1924.5 | 68.8 | 527 | 24 | ABU65889 | Human secreted |
| 20 | 1924.5 | 68.8 | 527 | 24 | ABU67666 | Human secreted/tira |
| 21 | 1924.5 | 68.8 | 527 | 24 | ABU65524 | Human PRO polypept |
| 22 | 1924.5 | 68.8 | 527 | 24 | ABU58660 | Human PRO polypept |
| 23 | 1924.5 | 68.8 | 527 | 24 | ABU56196 | Human secreted/tira |
| 24 | 1924.5 | 68.8 | 527 | 24 | ABU57191 | Human PRO polypept |
| 25 | 1924.5 | 68.8 | 527 | 24 | ABU10770 | Human secreted/tira |
| 26 | 1919.5 | 68.6 | 527 | 23 | ABP71237 | Human 32624 polypep |
| 27 | 1913.5 | 68.4 | 527 | 24 | ABU54566 | Human NOVX polypep |
| 28 | 1912 | 68.3 | 507 | 23 | AAE16940 | UDP-glucoronosyl c |
| 29 | 1874.5 | 67.0 | 454 | 23 | AAE26341 | Human UDP-glucosyl |
| 30 | 1795.5 | 64.2 | 548 | 22 | ABG27193 | Novel human diagno |
| 31 | 1699 | 60.7 | 533 | 23 | ABB98414 | Human NOV13, glucu |
| 32 | 1137.5 | 40.7 | 533 | 23 | AAU10940 | UDP glycosyltransf |
| 33 | 1137.5 | 40.7 | 533 | 23 | AAU10941 | UDP glycosyltransf |
| 34 | 1137.5 | 40.7 | 533 | 24 | ABB99721 | Amino acid sequenc |
| 35 | 1120.5 | 40.0 | 530 | 23 | ABB08490 | Human UGT1A9 amino |
| 36 | 1120.5 | 40.0 | 530 | 23 | ABB08491 | Human UGT1A9 amino |
| 37 | 1120.5 | 40.0 | 530 | 23 | ABB04429 | Human uridine 5'di |
| 38 | 1120.5 | 40.0 | 530 | 23 | ABB04430 | Human uridine 5'di |
| 39 | 1120.5 | 40.0 | 533 | 13 | AA26153 | HUG-Brl. Homo sap |
| 40 | 1114 | 39.8 | 544 | 12 | AA26154 | HUG-Brl. Homo sap |
| 41 | 1091 | 39.0 | 544 | 22 | ABG27065 | Novel human diagno |
| 42 | 1059 | 37.8 | 801 | 22 | ABG27195 | Novel human diagno |
| 43 | 803 | 28.7 | 541 | 23 | ABB57264 | Mouse ischaemic co |
| 44 | 765.5 | 27.4 | 245 | 21 | AAV57100 | UDP-glucuronosyltr |
| 45 | 695 | 24.8 | 516 | 22 | ABB62563 | Drosophila melanog |

ALIGNMENTS

| | | |
|----------|--|----------------------------|
| RESULT 1 | AA28677 | standard; Protein; 529 AA. |
| ID | AA28677 | |
| AC | AA28677 | |
| DT | 13-FEB-2001 | (first entry) |
| XX | | |
| DE | Human carbohydrate-modifying enzyme Incyte ID No: 2912330CD1. | |
| XX | | |
| KW | Human, carbohydrate-modifying enzyme; CME; antidiabetic; | |
| KW | immunosuppressive; anti-HIV; antiinflammatory; antianaemic; | |
| KW | antiatachmatic; antiarteriosclerotic; antihypoid; hepatocytic; | |
| KW | nephrotropic; antilgout; thymimetic; neuroprotective; osteopthtic; | |
| KW | antiarthritic; antiporiatic; uropathic; ophthalmological; | |
| KW | dermatological; antilucer; cytostatic; vitucide; antibacterial; | |
| KW | fungicide; protozoacide; tranquliser; vulnery; diabetes; | |
| KW | autoimmune disorder; inflammatory disorder; infection. | |
| XX | | |
| OS | Homo sapiens. | |
| XX | | |
| PN | W020006351-A2. | |
| XX | | |
| PD | 26-OCT-2000. | |
| XX | | |
| PF | 20-APR-2000; 2000WO-US10882. | |
| XX | | |
| PR | 21-APR-1999; 99US-0130383. | |
| XX | | |
| PA | (INCY-) INCYTE GENOMICS INC. | |
| XX | | |
| PI | Lal P, Yue H, Tang YT, Hillman JL, Baughn MR, Yang J; | |

XX WPI: 2000-672729/65.
 DR N-PSDB; AAC65396.
 XX Novel carbohydrate modifying enzyme polypeptides and polynucleotides
 PT for diagnosis, treatment, and prevention of carbohydrate metabolism
 PT disorders, autoimmune/inflammatory disorders, and cancer
 XX
 PS Claim 1; Page 71-72; 75pp; English.
 XX
 CC The present sequence is a human carbohydrate-modifying enzyme
 CC (CME). CME polynucleotides and polypeptides are useful for treating and
 CC diagnosing diseases associated with CME such as diabetes,
 CC autoimmune/inflammatory disorders such as AIDS, Addison's disease,
 CC adult respiratory distress syndrome, allergies, anaemia, asthma,
 CC atherosclerosis, autoimmune thyroiditis, bronchitis, cholecystitis,
 CC contact dermatitis, Crohn's disease, emphysema, erythroblastosis fetalis,
 CC glomerulonephritis, Good pasture's syndrome, gout, Grave's disease,
 CC Hashimoto's thyroiditis, multiple sclerosis, myasthenia gravis,
 CC osteoarthritis, osteoporosis, pancreatitis, polyomyelitis, psoriasis,
 CC Reiter's syndrome, arthritis, scleroderma, Sjogren's syndrome, systemic
 CC lupus erythematosus, ulcerative colitis, uveitis, Werner syndrome,
 CC complications of cancer, haemodialysis, and extracorporeal circulation,
 CC viral, bacterial, fungal parasitic, protozoal, and helminthic infections,
 CC trauma, or cancer. CME, or its catalytic or immunogenic fragment, is
 CC useful for drug screening.
 XX
 SQ Sequence 529 AA;

Query Match 100.0%; Score 2798; DB 21; Length 529;
 Best Local Similarity 100.0%; Pred. No. 5e-265;
 Matches 529; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSMKWTSLALLIQLSCYSSGCGKVLVWPTFESHMNIKITLDELVORGEHVTLASSA 60
 DB 1 MSMKWTSLALLIQLSCYSSGCGKVLVWPTFESHMNIKITLDELVORGEHVTLASSA 60
 QY SISPDNPSTLKEFVYVSVLTKTEFEDIKQLVKRMALPKDPTWVSFVSQVQEIIMTFN 120
 DB SISPDNPSTLKEFVYVSVLTKTEFEDIKQLVKRMALPKDPTWVSFVSQVQEIIMTFN 120
 QY 61 SISPDNPSTLKEFVYVSVLTKTEFEDIKQLVKRMALPKDPTWVSFVSQVQEIIMTFN 120
 DB 61 SISPDNPSTLKEFVYVSVLTKTEFEDIKQLVKRMALPKDPTWVSFVSQVQEIIMTFN 120
 QY 121 DILRFKCDIYNNKLMKKLOESRPDVLADAVPFGELALBLKIPVYSLRSPGYAI 180
 DB 121 DILRFKCDIYNNKLMKKLOESRPDVLADAVPFGELALBLKIPVYSLRSPGYAI 180
 QY 121 DILRFKCDIYNNKLMKKLOESRPDVLADAVPFGELALBLKIPVYSLRSPGYAI 180
 DB 121 DILRFKCDIYNNKLMKKLOESRPDVLADAVPFGELALBLKIPVYSLRSPGYAI 180
 QY 181 EKHSGGLFPSPYVNVVSELSQMTFIERVKNMIVLYFEFWFOIFDMKKWDQFYSVL 240
 DB 181 EKHSGGLFPSPYVNVVSELSQMTFIERVKNMIVLYFEFWFOIFDMKKWDQFYSVL 240
 QY 241 GRPPTLSEFMKADLWLRNWDPOFPHPLPNVEFGVGLHCKPKPKPKPKMESEEVQSSG 300
 DB 241 GRPPTLSEFMKADLWLRNWDPOFPHPLPNVEFGVGLHCKPKPKPKPKMESEEVQSSG 300
 QY 241 GRPPTLSEFMKADLWLRNWDPOFPHPLPNVEFGVGLHCKPKPKPKPKMESEEVQSSG 300
 DB 241 GRPPTLSEFMKADLWLRNWDPOFPHPLPNVEFGVGLHCKPKPKPKPKMESEEVQSSG 300
 QY 301 ENGUVVFSLSGWSVNTSEBRANVIASALAKIPQVLMRFPGDKPDTLGLNTRLYKMIQN 360
 DB 301 ENGUVVFSLSGWSVNTSEBRANVIASALAKIPQVLMRFPGDKPDTLGLNTRLYKMIQN 360
 QY 361 DILGHPKTKAFITTHGANGIYEAIVHGVPMGVPIFGDGLDNIAMKKAAGAAVEINFKTM 420
 DB 361 DILGHPKTKAFITTHGANGIYEAIVHGVPMGVPIFGDGLDNIAMKKAAGAAVEINFKTM 420
 QY 421 TSEDLLRLARTVITSSYKENAMRLSRHHDQVPLDPAVWIFPMWHRKAKGLRSAA 480
 DB 421 TSEDLLRLARTVITSSYKENAMRLSRHHDQVPLDPAVWIFPMWHRKAKGLRSAA 480
 QY 481 HDLTFQHSIDVIGFLTCVATAIFLFTKCLFSCQKNTKRIKRE 529
 DB 481 HDLTFQHSIDVIGFLTCVATAIFLFTKCLFSCQKNTKRIKRE 529

RESULT 2
 AAY78933
 standard; Protein; 528 AA.

XX AAY78933;
 AC 05-JUN-2000 (first entry)
 DT
 XX Human UDP-glucuronosyltransferase 2B4 amino acid sequence.
 DE
 XX Human UDP-glucuronosyltransferase 2B4; UGT2B4; polymorphism; metabolism; SNPs;
 XX drug interaction; detect; human; single nucleotide polymorphism.
 KM
 XX Homo sapiens.
 OS
 XX MO200006776-A1.
 PN
 XX 10-FEB-2000.
 PD
 XX 22-JUL-1999; 99MO-US16675.
 PF
 XX 28-JUL-1998; 98US-0094391.
 PR
 XX (AXYS-) AXYS PHARM INC.
 PA
 XX Galvin M, Miller A, Penny L, Riedy M;
 XX WPI: 2000-195321/17.
 DR N-PSDB; AA295199.
 XX
 PT Novel human UDP-glucuronosyltransferase sequence, polymorphisms for
 PT genotyping individuals to predict rate of metabolism of substrates and
 PT for identifying potential drug interactions
 XX
 PS Disclosure; Page 36-37; 72pp; English.
 XX

CC This sequence represents the human UDP-glucuronosyltransferase 2B4
 CC (UGT2B4) amino acid sequence. UDP-glucuronosyltransferase (UGTs) are a
 CC family of enzymes that catalyze the glucuronic acid conjugation of a
 CC wide range of endogenous and exogenous substrates. The UGT2B gene
 CC subfamily encode steroid metabolizing isoforms in the liver. Alteration
 CC of the expression or function of UGTs may effect drug metabolism. The
 CC invention relates to non-chromosomal nucleic acid molecules, which
 CC comprise human UGT2B sequence polymorphisms (see AA295051-295110). Probes
 CC which detect the UGT2B locus polymorphisms can be used to detect altered
 CC UGT2B metabolism of a substrate in an individual. The nucleic acid
 CC molecules comprising a human UGT2B sequence polymorphism can be used in
 CC screening assays for genotyping individuals, also to predict their rate
 CC of metabolism of UGT2B substrate, potential drug-drug interactions and
 CC adverse side effects. The polymorphisms can be used as single nucleotide
 CC polymorphisms (SNPs) for detecting genetic linkage related to phenotypic
 CC variation in activity or expression of UGT2B protein. The polymorphism
 CC containing nucleic acid molecules may also be used for generating
 CC genetically modified non-human animals and for obtaining site specific
 CC gene modification in cell lines.
 XX

Query Match 93.0%; Score 2602.5; DB 21; Length 528;
 Best Local Similarity 93.0%; Pred. No. 7.4e-246;
 Matches 492; Conservative 14; Mismatches 22; Indels 1; Gaps 1;

QY 1 MSMKWTSLALLIQLSCYSSGCGKVLVWPTFESHMNIKITLDELVORGEHVTLASSA 60
 DB 1 MSMKWTSLALLIQLSCYSSGCGKVLVWPTFESHMNIKITLDELVORGEHVTLASSA 60
 QY 61 SISPDNPSTLKEFVYVSVLTKTEFEDIKQLVKRMALPKDPTWVSFVSQVQEIIMTFN 120
 DB 61 SISPDNPSTLKEFVYVSVLTKTEFEDIKQLVKRMALPKDPTWVSFVSQVQEIIMTFN 120
 QY 121 DILRFKCDIYNNKLMKKLOESRPDVLADAVPFGELALBLKIPVYSLRSPGYAI 180
 DB 121 DILRFKCDIYNNKLMKKLOESRPDVLADAVPFGELALBLKIPVYSLRSPGYAI 180
 QY 181 EKHSGGLFPSPYVNVVSELSQMTFIERVKNMIVLYFEFWFOIFDMKKWDQFYSVL 240
 DB 181 EKHSGGLFPSPYVNVVSELSQMTFIERVKNMIVLYFEFWFOIFDMKKWDQFYSVL 240

Not applicable

Db 181 EKHSGLLPSPSYVPMVMSLSQMTFIERVKMIYLYFEFFWQIIPDMKKMOQFYSEVL 240
 Qy 241 GRPTLSETMAKADIWILIRNYWDPQFPHLLPNVEFVGLHCCKPAKPLPKMEEFVQSSG 300
 Db 241 GRPTLSETMAKADIWILIRNYWDPQFPHLLPNVEFVGLHCCKPAKPLPKMEEFVQSSG 300
 Qy 301 ENGVAVPSLGSVMNTSEERANVIASALAKI POKVLMRFQDNKPDITGLNTRYKWIIPON 360
 Db 301 ENGVAVPSLGSVMNTSEERANVIASALAKI POKVLMRFQDNKPDITGLNTRYKWIIPON 360
 Qy 361 DLGHPRKTAFTTHGANGNGIYEALYHGVPMVGPVIFGQDLNIAHMKAKGAIVEINFTKM 420
 Db 361 DLGHPRKTAFTTHGANGNGIYEALYHGVPMVGPVIFGQDLNIAHMKAKGAIVEINFTKM 420
 Qy 421 TSEDLRALRTVITDSSYKENAMRLSRIRHDPVKPLDRAVFWIEFWMRHKGAKHLRSAA 480
 Db 421 TSEDLRALRTVITDSSYKENAMRLSRIRHDPVKPLDRAVFWIEFWMRHKGAKHLRSAA 480
 Qy 481 HDLTFQHYSDIVIGFLTTCVATAIFLTCKFLFSCQKFNKTRKIEKRE 529
 Db 481 HDLTFQHYSLDVTGFLACVATVIFITTC-LFCVMKFRVTRGKKGRD 528
 RESULT 3
 ABG05525 ID ABG05525 standard; Protein; 540 AA.
 XX AC ABG05525;
 XX DT 13-FEB-2002 (first entry)
 XX DE Novel human diagnostic protein #5516.
 XX KM Human: chromosome mapping; gene mapping; gene therapy; forensic;
 KM food supplement; medical imaging; diagnostic; genetic disorder.
 XX OS Homo sapiens.
 XX PN MO200175067-A2.
 XX PD 11-OCT-2001.
 XX PF 30-MAR-2001; 2001WO-US08631.
 XX PR 31-MAR-2000; 2000US-0540217.
 XX PR 23-AUG-2000; 2000US-0649167.
 XX PA (HYSE-) HYSEQ INC.
 XX PI Drmanac RT, Liu C, Tang YT;
 XX DR WPI, 2001-639362/73.
 XX DR N-PSDB; AAS69712.
 XX PT New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity -
 PS Claim 20; SEQ ID No 35884; 103bp; English.
 XX CC The invention relates to isolated polynucleotide (I) and
 CC polypeptide (II) sequences. (I) is useful as hybridization probes,
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
 CC and gene mapping, and in recombinant production of (II). The
 CC polynucleotides are also used in diagnostics as expressed sequence tags
 CC for identifying expressed genes. (I) is useful in gene therapy techniques
 CC to restore normal activity of (II) or to treat disease states involving
 CC (II). (II) is useful for generating antibodies against it, detecting or
 CC quantitating a polypeptide in tissue, as molecular weight markers and as
 CC a food supplement. (II) and its binding partners are useful in medical
 CC imaging of sites expressing (II). (I) and (II) are useful for treating
 CC disorders involving aberrant protein expression or biological activity.

CC The polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. ABG00010-ABG03077 represent novel human
 CC diagnostic amino acid sequences of the invention.
 CC Note: The sequence data for this patent did not appear in the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 540 AA;
 Query Match 92.4%; Score 2584.5; DB 22; Length 540;
 Best Local Similarity 92.4%; Pred. No. 4,4e-244;
 Matches 489; Conservative 14; Mismatches 25; Indels 1; Gaps 1;
 Qy 1 MSKMWTSAALLLIQSLSCYFSSGSGKVLWPTFESHMNNITILDELVQRGHEVTVLASSA 60
 Db 13 MSKMWTSAALLLIQSLSCYFSSGSGKVLWPTFESHMNNITILDELVQRGHEVTVLASSA 72
 Qy 61 SISFDPSPTLKREVPVSLTKTEFEDITKOLYKRAELPKCTFWSYFQVOEIMWTFN 120
 Db 73 SISFDPSPTLKREVPVSLTKTEFEDITKOLYKRAELPKCTFWSYFQVOEIMWTFN 132
 Qy 121 DILRKPCDIVSNKKLMKKLQESRFVDVVLADAVPFEGELAEELKIPFVYSLFSPGYAI 180
 Db 133 DILRKPCDIVSNKKLMKKLQESRFVDVVLADAVPFEGELAEELKIPFVYSLFSPGYAI 192
 Qy 181 EKHSGLLPSPSYVPMVMSLSQMTFIERVKMIYLYFEFFWQIIPDMKKMOQFYSEVL 240
 Db 193 EKHSGLLPSPSYVPMVMSLSQMTFIERVKMIYLYFEFFWQIIPDMKKMOQFYSEVL 252
 Qy 241 GRPTLSETMAKADIWILIRNYWDPQFPHLLPNVEFVGLHCCKPAKPLPKMEEFVQSSG 300
 Db 253 GRPTLSETMAKADIWILIRNYWDPQFPHLLPNVEFVGLHCCKPAKPLPKMEEFVQSSG 312
 Qy 301 ENGVAVPSLGSVMNTSEERANVIASALAKI POKVLMRFQDNKPDITGLNTRYKWIIPON 360
 Db 313 ENGVAVPSLGSVMNTSEERANVIASALAKI POKVLMRFQDNKPDITGLNTRYKWIIPON 372
 Qy 361 DLGHPRKTAFTTHGANGNGIYEALYHGVPMVGPVIFGQDLNIAHMKAKGAIVEINFTKM 420
 Db 373 DLGHPRKTAFTTHGANGNGIYEALYHGVPMVGPVIFGQDLNIAHMKAKGAIVEINFTKM 432
 Qy 421 TSEDLRALRTVITDSSYKENAMRLSRIRHDPVKPLDRAVFWIEFWMRHKGAKHLRSAA 480
 Db 433 TSEDLRALRTVITDSSYKENAMRLSRIRHDPVKPLDRAVFWIEFWMRHKGAKHLRSAA 492
 Qy 481 HDLTFQHYSDIVIGFLTTCVATAIFLTCKFLFSCQKFNKTRKIEKRE 529
 Db 493 HDLTFQHYSLDVTGFLACVATVIFITTC-LFCVMKFRVTRGKKGRD 540
 RESULT 4
 AAO22794 ID AAO22794 standard; Protein; 529 AA.
 XX AC AAO22794;
 XX DT 21-NOV-2002 (first entry)
 XX DE Protein of drug metabolizing enzyme - 7493833CD1.
 XX KM Anti-HIV, antiarteriosclerotic; dermatological; cytostatic; thyromimetic;
 KM osteopathic; thrombolytic; ophthalmological; antidiabetic; hepatotropic;
 KM antidiarrheic; antiinflammatory; virocidic; immunogenic; autoimmune;
 KM drug metabolizing enzyme; DME; inflammatory; AIDS; atherosclerosis;
 KM contact dermatitis; cell-proliferative; cancer; cirrhosis; dwarfism;
 KM developmental; hypothyroidism; endocrine; osteoporosis; thrombosis;
 KM diabetes; glaucoma; keratitis; metabolic; hyperlipidaemia; diarrhoea;
 KM cystic fibrosis; gastrointestinal; gastroenteritis; liver; hepatitis;
 KM Reye's syndrome; exogenous compound; gene therapy; enzyme; human.

OS Homo sapiens.
 XX
 XX WO200266654-A2.
 XX
 XX 29-AUG-2002.
 PD
 XX 14-FEB-2002; 2002WO-US04918.
 PF
 XX 16-FEB-2001; 2001US-269643P.
 PR 23-FEB-2001; 2001US-271332P.
 PR 16-MAR-2001; 2001US-276767P.
 PR 06-APR-2001; 2001US-282077P.
 PR 19-APR-2001; 2001US-285447P.
 PR 27-APR-2001; 2001US-287060P.
 PR 03-MAY-2001; 2001US-288543P.
 PR
 XX (INCY-) INCYTE GENOMICS INC.
 PA
 XX Astronomoff A, Au-young J, Baughn MR, Ding L, Duggan BW;
 PI Forsythe J, Gietzen KJ, Griffin JB, Lee EA, Lu Y, Richardson TW;
 PI Ring HZ, Sanjanna MM, Swarnakar A, Walla NK, Warren BA, Xu Y;
 PI Yue H, Zabarjadian Y;
 XX
 XX WPI; 2002-674949/72.
 DR N-PSDB; AAL41490.
 DR
 XX New drug metabolizing enzymes (DME) useful for diagnosing, treating and
 PT preventing diseases or conditions associated with aberrant DME
 PT expression, e.g. cancer, AIDS, atherosclerosis, diabetes, glaucoma,
 PT hepatitis, osteoporosis -
 XX
 PS Claim 1; Page 150-151; 166pp; English.
 XX
 CC The invention relates to an isolated polypeptide comprising 12 fully
 CC defined sequences of 81-615 amino acids given in the specification; a
 CC naturally occurring amino acid sequence at least 90% identical to, having
 CC 81-599 amino acids, at least 93% identical to a sequence of 529 amino
 CC acids, or at least 97% identical to a sequence of 615 amino acids, all
 CC given in the specification; or a biologically active or immunogenic
 CC fragment of the polypeptide. The polypeptides and polynucleotides are
 CC useful in diagnosing, treating and preventing diseases or conditions
 CC associated with the decreased expression or overexpression of a drug
 CC metabolizing enzyme (DME), such as autoimmune/inflammatory (e.g. AIDS,
 CC atherosclerosis, contact dermatitis) cell-proliferative (e.g. cancer,
 CC cirrhosis), developmental (e.g. dwarfism, hypothyroidism), endocrine
 CC (e.g. osteoporosis, chromoblastoma, diabetes), eye (e.g. glaucoma,
 CC keratitis), metabolic (e.g. hyperlipidaemia, cystic fibrosis),
 CC gastrointestinal (e.g. gastroenteritis, diarrhoea), or liver (e.g.
 CC hepatitis, Reye's syndrome) disorders. These are also useful in assessing
 CC the effects of exogenous compounds on the expression of nucleic acid and
 CC amino acid sequences of DME. The DME or its fragments are useful in
 CC screening compounds for effectiveness as agonist or antagonist of the
 CC polypeptides, or in altering the expression of the target polynucleotide
 CC and compounds that specifically bind to or modulate the activity of the
 CC polypeptide. The microarray is useful in monitoring or measuring protein-
 CC protein interactions, drug-target interactions, and gene expression
 CC profiles. The polynucleotides of the invention can be used in gene
 CC therapy. This sequence represents a drug metabolizing enzyme protein of
 CC the invention.
 CC
 XX
 XX Sequence 529 AA:
 SQ
 Query Match 89.5%; Score 2503; DB 23; Length 529;
 Best Local Similarity 87.7%; Pred. No. 4.2e-236;
 Matches 464; Conservative 29; Mismatches 36; Indels 0; Gaps 0;

QY 121 DILKPKCDIVSNKLMKKLQESRFDVVLADAVFPFGELLAELKIPFVYSLNPSGYAI 180
 |||
 DB 121 DILKPKCDIVSNKLMKKLQESRFDVLPADALFPCSELLAELFNIPFVYSLNPSGYAI 180
 |||
 QY 181 EKHSGLLPFPSPYVPMVMSLSDQMTFIEKVKNNIYVLYEFMFQIFDMKKMDQFYSEVL 240
 |||
 DB 181 EKHSGLFPFPSPYVPMVMSLSDQMTFIEKVKNNIYVLYEFMFQIFDMKKMDQFYSEVL 240
 |||
 QY 241 GRPPTLSEMAKDIIWIRVYMDPQPHPLPVPVEFGJHCPKAPKPKMESEFVSSG 300
 |||
 DB 241 GRPPTLSEMAKDIIWIRVYMDPQPHPLPVPVEFGJHCPKAPKPKMESEFVSSG 300
 |||
 QY 301 ENGIVVPSLGSMTSNTSEERANVYASALAKIPKVLMPFGNKPDTGLNTRLYKWIPON 360
 |||
 DB 301 ENGIVVPSLGSMTSNTSEERANVYASALAKIPKVLMPFGNKPDTGLNTRLYKWIPON 360
 |||
 QY 361 DLIGHPTKAFITHGNGNGIYEALYHGVPMVGVPIFGDQDNLIAHMKAKAAVEINFTKM 420
 |||
 DB 361 DLIGHPTKAFITHGNGNGIYEALYHGVPMVGVPIFGDQDNLIAHMKAKAAVEINFTKM 420
 |||
 QY 421 TSEDLALALRTVITDSSYKENAMKLSRIHHDQPKPLDPAVFWTEFVRRHKGAKHLSAA 480
 |||
 DB 421 TSEDLALALRTVITDSSYKENAMKLSRIHHDQPKPLDPAVFWTEFVRRHKGAKHLSAA 480
 |||
 QY 481 HDLTFQHYSDIVGFLITCVATLFFPTKCFPLSCQKFNTRKIEKRE 529
 |||
 DB 481 HDLTFQHYSDIVGFLITCVATLFFPTKCFPLSCQKFNTRKIEKRE 529
 |||
 RESULT 5
 AAE28617
 ID AAE28617 standard; Protein; 529 AA.
 AC AAE28617;
 XX
 DT 27-DEC-2002 (first entry)
 XX
 XX Human UGT2B7 protein.
 DE
 XX Human; UDP-glucuronosyl transferase; UGT; UGT2B7; toxicity; cancer;
 KW therapy; epirubicin; cytosol; enzyme.
 XX
 OS Homo sapiens.
 XX
 XX WO200259375-A2.
 XX
 XX 01-AUG-2002.
 XX
 XX 25-JAN-2002; 2002WO-US02083.
 XX
 XX 26-JAN-2001; 2001US-264534P.
 XX
 XX (UYCH-) UNIV CHICAGO.
 PA
 XX Ratain MJ, Innocenti F, Das S, Iyer L, Sawyer M;
 PI WPI; 2002-691534/74.
 DR N-PSDB; AAD45991.
 DR
 XX Determining the dose of a UGT2B7-glucuronidated drug for treating
 PT cancer, comprises determining the level of UGT2B7 activity or
 PT expression in a patient -
 XX
 PS Disclosure; Page 144-145; 160pp; English.
 XX
 CC The invention relates to an UDP-glucuronosyl transferase (UGT) enzyme,
 CC UGT2B7. The invention also relates to compositions and methods for
 CC optimizing UGT2B7 substrate dosings and for predicting UGT2B7 substrate
 CC toxicity. The method is useful in determining the dose of a UGT2B7-
 CC glucuronidated drug that may be used in treating cancer patients. It
 CC is also useful in determining persons at risk for epirubicin toxicity,
 CC in reducing or eliminating side effects associated with epirubicin

CC treatment, and in ways of increasing the efficacy of dosage regimens.
CC The present sequence is human UG72B7 protein.

XX Sequence 529 AA;

Query Match 83.3%; Score 2330; DB 23; Length 529;
Best Local Similarity 81.1%; Pred. No. 3.9e-219;
Matches 429; Conservative 42; Mismatches 58; Indels 0; Gaps 0;

```
QY 1 MSMTATALLILQISCFSSGCGKVLVMPTEFSHMNNTITIDELVORGEVTLASSA 60
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
DB 1 MSVMTATVILTIQISFCFSSGCGKVLVMAEYSHMNMKITIDELVORGEVTLASSA 60
QY 61 SIFDPPSPSTLKEVYPVSLTKTEFEDIKQVYKMAELPKOTFMSYFSQVOEIMWTFN 120
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
DB 61 SILFDPNNSALKIEIYPTSLTKLEENFMQIKRNSDLPKOTFMYLFSQVOEIMWTFN 120
QY 121 DILRKFCQDIVSNKKMKKLQESRPDVVLADAVFPFGEELAEELKIPVYSLRFSGYAI 180
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
DB 121 DILRKFCQDIVSNKKMKKLQESRPDVVLADAVFPFGEELAEELKIPVYSLRFSGYAI 180
QY 181 EKHSGLLPSPSYVPVMSSELSDQMTFIERVKMNIYVLFYEFMFOIDMKKMDQFYSEVL 240
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
DB 181 EKHSGLLPSPSYVPVMSSELSDQMTFIERVKMNIYVLFYEFMFOIDMKKMDQFYSEVL 240
QY 241 GRPTTSETMAKADIMILIRNYMDFOPHPILPNVEFGGLHCKPAKPLPKEMEFVQSSG 300
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
DB 241 GRPTTSETMAKADIMILIRNSMNFQFHPILPNVDFGGLHCKPAKPLPKEMEFVQSSG 300
QY 301 ENGVVVFSLGSVNSNTSEERANVIASALAKIPQKVLMPDGNKPDITGLNTRYKWIPON 360
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
DB 301 ENGVVVFSLGSVNSNTSEERANVIASALAKIPQKVLMPDGNKPDITGLNTRYKWIPON 360
QY 361 DLIGHPKTKAFITGGNNGIYEALTYHGVPMGVPIFGDQDNIAMKAAVEINIKTM 420
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
DB 361 DLIGHPKTKAFITGGANGIYEALTYHGIPIWGIPLFADQPDNIAMKAAVAVDNFTM 420
QY 421 TSEDLRALRTVITDSSYKENAMRLSRIHDPVKPLDRAVFIETFWMRKHGALHSAA 480
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
DB 421 SSTDLNALRKVINDPYKENVKLSRIQHDQVPKPLDRAVFIETFWMRKHGALHSAA 480
QY 481 HDLTWFQHSYIDVIGFLITCVATAIFLFTKCFLFSCKQFNKTRKIEKRE 529
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
DB 481 HDLTWFQHSYIDVIGFLITCVATAIFLFTKCFLFSCKQFNKTRKIEKRE 529
```

RESULT 6
ABG05523
ID ABG05523 standard; Protein; 533 AA.

XX ABG05523;

DT 13-FEB-2002 (first entry)

DE Novel human diagnostic protein #5514.

KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.

OS Homo sapiens.

PN WO200175067-A2.

PD 11-OCT-2001.

PF 30-MAR-2001; 2001WO-US08631.

PR 31-MAR-2000; 2000US-0540217.

PR 23-AUG-2000; 2000US-0649167.

PA (HYSE-) HYSEO INC.

PI Drmanac RT, Liu C, Tang YT;

XX WI: 2001-639362/73.
DR N-PSDB; AAS69710.

PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensic, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -

XX Claim 20; SEQ ID No 35882; 103pp; English.

XX The invention relates to isolated polynucleotide (I) and
XX polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG00010-ABG30377 represent novel human
CC diagnostic amino acid sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 533 AA;

Query Match 83.1%; Score 2325; DB 22; Length 533;
Best Local Similarity 80.9%; Pred. No. 1.2e-218;
Matches 428; Conservative 43; Mismatches 58; Indels 0; Gaps 0;

```
QY 1 MSMTATALLILQISCFSSGCGKVLVMPTEFSHMNNTITIDELVORGEVTLASSA 60
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
DB 5 MSVMTATVILTIQISFCFSSGCGKVLVMAEYSHMNMKITIDELVORGEVTLASSA 64
QY 61 SIFDPPSPSTLKEVYPVSLTKTEFEDIKQVYKMAELPKOTFMSYFSQVOEIMWTFN 120
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
DB 65 SILFDPNNSALKIEIYPTSLTKLEENFMQIKRNSDLPKOTFMYLFSQVOEIMWTFN 124
QY 121 DILRKFCQDIVSNKKMKKLQESRPDVVLADAVFPFGEELAEELKIPVYSLRFSGYAI 180
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
DB 125 DILRKFCQDIVSNKKMKKLQESRPDVVLADAVFPFGEELAEELKIPVYSLRFSGYAI 184
QY 181 EKHSGLLPSPSYVPVMSSELSDQMTFIERVKMNIYVLFYEFMFOIDMKKMDQFYSEVL 240
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
DB 185 EKHSGLLPSPSYVPVMSSELSDQMTFIERVKMNIYVLFYEFMFOIDMKKMDQFYSEVL 244
QY 241 GRPTTSETMAKADIMILIRNYMDFOPHPILPNVEFGGLHCKPAKPLPKEMEFVQSSG 300
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
DB 245 GRPTTSETMAKADIMILIRNSMNFQFHPILPNVDFGGLHCKPAKPLPKEMEFVQSSG 304
QY 301 ENGVVVFSLGSVNSNTSEERANVIASALAKIPQKVLMPDGNKPDITGLNTRYKWIPON 360
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
DB 305 ENGVVVFSLGSVNSNTSEERANVIASALAKIPQKVLMPDGNKPDITGLNTRYKWIPON 364
QY 361 DLIGHPKTKAFITGGNNGIYEALTYHGVPMGVPIFGDQDNIAMKAAVEINIKTM 420
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
DB 365 DLIGHPKTKAFITGGANGIYEALTYHGIPIWGIPLFADQPDNIAMKAAVAVDNFTM 424
QY 421 TSEDLRALRTVITDSSYKENAMRLSRIHDPVKPLDRAVFIETFWMRKHGALHSAA 480
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
DB 425 SSTDLNALRKVINDPYKENVKLSRIQHDQVPKPLDRAVFIETFWMRKHGALHSAA 484
QY 481 HDLTWFQHSYIDVIGFLITCVATAIFLFTKCFLFSCKQFNKTRKIEKRE 529
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
```

Db 485 HDLTFQYHSLDVIGFLVCAATVIFVTCKCLFCFWMKARKAKKND 533

RESULT 7
AA78934
ID AA78934 standard; Protein; 524 AA.
XX
AC AA78934;
XX
DT 05-JUN-2000 (first entry)
XX
DE Human UDP-glucuronosyltransferase 2B7 amino acid sequence.
XX
KW UDP-glucuronosyltransferase 2B4; UGT2B4; polymorphism; metabolism; SNPs;
KM drug interaction; detect; human; single nucleotide polymorphism.
XX
OS Homo sapiens.
XX
PN MO200006776-A1.
XX
PD 10-FEB-2000.
XX
PE 22-JUL-1999; 99WO-US16675.
XX
PR 28-JUL-1998; 98US-0094391.
XX
PA (AAYS-) AAYS PHARM INC.
XX
PI Galvin M, Miller A, Penny L, Riedy M;
XX
DR WPI: 2000-195321/17.
DR N-PSDB; AA295200.
XX
PT Novel human UDP-glucuronosyltransferase sequence, polymorphisms for
PT genotyping individuals to predict rate of metabolism of substrates and
PT for identifying potential drug interactions
XX
PS Disclosure: Page 44-45; 72pp; English.

XX This sequence represents the human UDP-glucuronosyltransferase 2B7
CC (UGT2B7) amino acid sequence. UDP-glucuronosyltransferase (UGTs) are a
CC family of enzymes that catalyse the glucuronic acid conjugation of a
CC wide range of endogenous and exogenous substrates. The UGT2B gene
CC subfamily encode stereoid metabolizing isoforms in the liver. Alteration
CC of the expression or function of UGTs may effect drug metabolism. The
CC invention relates to non-chromosomal nucleic acid molecules, which
CC comprise human UGT2B sequence polymorphisms (see AA295051-295110). Probes
CC which detect the UGT2B locus polymorphisms can be used to detect altered
CC UGT2B metabolism of a substrate in an individual. The nucleic acid
CC molecules comprising a human UGT2B sequence polymorphism can be used in
CC screening assays for genotyping individuals, also to predict their rate
CC of metabolism of UGT2B substrate, potential drug-drug interactions and
CC adverse side effects. The polymorphisms can be used as single nucleotide
CC polymorphisms (SNPs) for detecting genetic linkage related to phenotypic
CC variation in activity or expression of UGT2B protein. The polymorphism
CC containing nucleic acid molecules may also be used for generating
CC genetically modified non-human animals and for obtaining site specific
CC gene modification in cell lines.

XX
XX
SQ Sequence 524 AA;

Query Match 83.0%; Score 2322; DB 21; Length 524;
Best Local Similarity 81.5%; Pred. No. 2,3e-216;
Matches 427; Conservative 42; Mismatches 55; Indels 0; Gaps 0;

QY 1 MSMKWTSAALLIQLSCYFSSGSGCKVLVPTFSHMNKKITLDELVQGHVEVTLASSA 60
DB 1 MSVKRTSVILQLSCFSSGCKVLVAAEYSHMNKKITLDELVQGHVEVTLASSA 60
QY 61 SISFDPNSBTLKFEVYVSLKTEPEDIIKQLVKRAEALPQDTWMSYFSQVQEIIMWTFN 120
DB 61 SILFDPNNSALKIEIYPTSLKTELENFIMQOIKRWSDLPQDTFWLVSQVQEIIMSIIFG 120

QY 121 DLRKPKCDIVSNKKMLKQLQESRFDVVLADAVFPFGELLAEELKIPVYSLRSPGYAI 180
DB 121 DTRKPKCDIVSNKKMLKQLQESRFDVIFADAFPCSELLAELENIPVYSLRSPGYTE 180
QY 181 EKISGGLFPSPYVPVMSLSDOMTFIERVKMIIYLYEFEPQIIPDMKKMQOFSYEV 240
DB 181 EKISGGFIFFPSYVPVMSLSLTDQMTFMRVKMIIYLYEFEPQIIPDMKKMQOFSYEV 240
QY 241 GRPTLSEITAKADIVLIRVWDPQFPHPPLPNVEFVGLHCKRAKPLPKMEEFVQSSG 300
DB 241 GRPTLSEITAKADIVLIRVWDPQFPHPPLPNVEFVGLHCKRAKPLPKMEEFVQSSG 300
QY 301 ENGIVVPSLGSWVSNTESEERANVIALAKIPQVLRFPDGNKPDYLGITRLYKWIPO 360
DB 301 ENGIVVPSLGSWVSNTESEERANVIALAKIPQVLRFPDGNKPDYLGITRLYKWIPO 360
QY 361 DLGHPEKTAFTITGGNNGIYEALYHGVNPGVIFPDQDNLNTHMKAKAGAAVINFKT 420
DB 361 DLGHPEKTAFTITGGNNGIYEALYHGVNPGVIFPDQDNLNTHMKAKAGAAVINFKT 420
QY 421 TSEDLRLALRTVITDSYKENAMRLSRHHDPYKPLDRAVFWIEFVRRHKGAGHLSAA 480
DB 421 SSTDLNALAKRVINDPSYKENMKLSRIQHDQPKPLDRAVFWIEFVRRHKGAGHLSAA 480
QY 481 HDLTFQYHSLDVIGFLVCAATVIFVTCKCLFCFWMKARKAK 524
DB 481 HDLTFQYHSLDVIGFLVCAATVIFVTCKCLFCFWMKARKAK 524

RESULT 8
AAE02188
ID AAE02188 standard; Protein; 529 AA.
XX
AC AAE02188;
XX
DT 06-AUG-2001 (first entry)
XX
DE Human breast cancer specific gene-2 (BCSG-2) protein.
XX
KW Human; breast cancer specific gene-2; BCSG-2; cytostatic; vaccine;
KM breast cancer; therapeutic; gene therapy.
XX
OS Homo sapiens.
XX
PN MO200137779-A2.
XX
PD 31-MAY-2001.
XX
PE 22-NOV-2000; 2000WO-US32056.
XX
PR 23-NOV-1999; 99US-0166973.
XX
PA (DIAD-) DIADEXUS INC.
XX
PI Salceda S, Caferkey R, Recipon H, Sun Y;
XX
DR WPI: 2001-367602/38.
DR N-PSDB; AAD06860.
XX
PT Novel breast cancer specific gene for diagnosing, monitoring, staging,
PT imaging, preventing and treating cancers, particularly breast cancer
XX
PS Claim 2; Page 55-57; 66pp; English.

XX The invention relates human breast cancer specific genes (BCSG's) and
CC their corresponding proteins. BCSG is useful for diagnosing, staging,
CC monitoring, imaging, preventing and treating breast cancers. BCSG is also
CC useful for inducing an immune response against a target cell expressing
CC BCSG. The invention also provide methods for detecting genetic lesions or
CC mutations in BCSG, thereby determining if a human with the genetic lesion
CC is at risk for breast cancer or has breast cancer. BCSG antibodies
CC labelled with paramagnetic ions or radioisotopes is useful for imaging
CC breast cancers, while BCSG antibodies conjugated to a cytotoxic agent is

CC useful for treating breast cancer. BCSG is useful in the rational design
CC of new therapeutics for imaging and treating cancers. BCSG is also used
CC in gene therapy. The present sequence is human breast cancer specific
CC gene-2 (BCSG-2) protein.

XX Sequence 529 AA;

Query Match 81.9%; Score 2292; DB 22; Length 529;
Best Local Similarity 80.2%; Pred. No. 2e-215;
Matches 424; Conservative 43; Mismatches 62; Indels 0; Gaps 0;

```
QY 1 MSKMTSALLLIQLSCYSSGCGKVLVPTFESHMNIKITIDELVORGEVTLASSA 60
DB 1 MTLKMTSVLLIHLISCYSSGCGKVLVMAEYSHMMNMKITLKLVELYORGEVTLASSA 60
QY 61 SISFDPNPSTLKEVYVPSLTKTEFEDIIKOLVYKMAELIKDTFWSFSQVOEIMTFEN 120
DB 61 SILFDPNASTLKEVYVPSLTKTEFENIMQOVYRMSDIKRSDFWLSQEOEILWELV 120
QY 121 DILKRFCKDIYSNKKLMKKLOESRPDVVLADAVFPFGLAELKIPVYSLRSPGYAI 180
DB 121 DIFNFCQDVSNKKMKKLOESRPDIYFADAVFPFGLAELALNIRVYSLRFTPGYTI 180
QY 181 EKHSGLIFPPSYVYVPMVSELSQMTFIERVKMNIYLYFEFPOIFDMKKMDQFYSEVL 240
DB 181 ERHSGGLIFPPSYPIVMSKLSQDQTFMERVYKMIYLYFPFWMQMSDKMDQFYSEVL 240
QY 241 GRPTLSTMAKADIVLIRNWDQFPHPLPNVEFYVGLHCKRAKPLPKMEEFVQSSG 300
DB 241 GRPTLLEPTMKADIMLRNWSQFPHPLPNVDFVGGFHCKRAKPLPKMEEFVQSSG 300
QY 301 ENGIVVFSLSGWSNTESEERANVIASALAKIPQKVLMPFGDNKPTGLNTRLYKMIPO 360
DB 301 ENGIVVFSLSGVISNMTERRANVATATAKIPQKVLMPFGDNKPDALNTRLYKMIPO 360
QY 361 DLGHPKTRAFITGGANGIYEAIYHGVPMVGPPIFGDQLDNIAMKAGAAVEINFXTM 420
DB 361 DLGHPKTRAFITGGANGIYEAIYHGPVIGPIFGDQDNIAHMKAGAAVRLDFXTM 420
QY 421 TSEDLRLALRTVITDSSYKENAMRLSRHHDPQVPLDRAVPMIEFVWRHKGAKHLRSA 480
DB 421 SSTDLNALKTIVINDPLKENIMKLSRIQHDQPVKPLDRAVPMIEFVWRHKGAKHLRSA 480
QY 481 HDLTFMFOHSIDVIGFLTCAVATPIFPTKCFLFSCKQFNTRKIEKRE 529
DB 481 HDLTFMFOHSLDVIGFLACVATVIFITKCFLFCFMKFAKKGKKRD 529
```

RESULT 9
ABG05524 ID ABG05524 standard; Protein: 532 AA.
XX AC ABG05524;
XX AC ABG05524;
DT 13-FEB-2002 (first entry)
XX DE Novel human diagnostic protein #5515.
XX KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX KW food supplement; medical imaging; diagnostic; genetic disorder.
XX OS Homo sapiens.
XX PN W0200175067-A2.
XX PD 11-OCT-2001.
XX PF 30-MAR-2001; 2001WO-US08631.
XX PR 31-MAR-2000; 2000US-0540217.
XX PR 23-AUG-2000; 2000US-0649167.
XX PA (HYSE-) HYSEQ INC.

XX XX
PI Drmanac RT, Liu C, Tang YT;
XX WPI: 2001-639362/73.
DR N-PSDB: AAS69711.

PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -

XX Claim 20; SEQ ID No 35883; 103pp; English.

PS The invention relates to isolated polynucleotide (I) and
XX polypeptide (II) sequences. (I) is useful as hybridisation probes,
XX polynucleotide chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG00010-ABG03037 represent novel human
CC diagnostic amino acid sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 532 AA;

Query Match 79.8%; Score 2233; DB 22; Length 532;
Best Local Similarity 79.4%; Pred. No. 1.3e-209;
Matches 421; Conservative 43; Mismatches 64; Indels 2; Gaps 2;

```
QY 1 MSKMTSALLLIQLSCYSSGCGKVLVPTFESHMNIKITIDELVORGEVTLASSA 60
DB 4 MALKMTT-VLLIQSFFSSGCGKVLVMAEYSHMMNMKITLKLVELYORGEVTLASSA 62
QY 61 SISFDPNPSTLKEVYVPSLTKTEFEDIIKOLVYKMAELIKDTFWSFSQVOEIMTFEN 120
DB 63 SILFDPNDSSTLKEVYVPSLTKTEFENIMQOVYKRLSEIQKDTFWMFQSOEILWAIN 122
QY 121 DILKRFCKDIYSNKKLMKKLOESRPDVVLADAVFPFGLAELKIPVYSLRSPGYAI 180
DB 123 DILKRFCKDVSNKKMKKLOESRPDIYFADAVLPFGLAELALNIRVYSLRFTPGYTI 182
QY 181 EKHSGLIFPPSYVYVPMVSELSQMTFIERVKMNIYLYFEFPOIFDMKKMDQFYSEVL 240
DB 183 ERHSGGLIFPPSYVYVPMVSKLSQDQTFMERVYKMIYLYFPFWMQMSDKMDQFYSEVL 242
QY 241 GRPTLSTMAKADIVLIRNWDQFPHPLPNVEFYVGLHCKRAKPLPKMEEFVQSSG 300
DB 243 GRPTLSETMKADIMLRNWSNMFKPHPLPNVDFVGGFHCKRAKPLPKMEEFVQSSG 302
QY 301 ENGIVVFSLSGWSNTESEERANVIASALAKIPQKVLMPFGDNKPTGLNTRLYKMIPO 359
DB 303 ENGIVVFSLSGWSNTESEERANVATATAKIPQKVLMPFGDQDNIAHMKAGAAVRLDFXTM 362
QY 360 NDLLGHPKTRAFITGGANGIYEAIYHGVPMVGPPIFGDQLDNIAMKAGAAVEINFXT 419
DB 363 NDLLGHPKTRAFITGGANGIYEAIYHGPVIGPIFGDQDNIAHMKAGAAVRLDFXTM 422
QY 420 MTSDDLRLALRTVITDSSYKENAMRLSRHHDPQVPLDRAVPMIEFVWRHKGAKHLRSA 479
DB 423 MTSDDLNALKTIVINDPLKENIMKLSRIQHDQPVKPLDRAVPMIEFVWRHKGAKHLRSA 482
```


CC polymorphisms (SNPs) for detecting genetic linkage related to phenotypic
 CC variation in activity or expression of UGT2B protein. The polymorphism
 CC containing nucleic acid molecules may also be used for generating
 CC genetically modified non-human animals and for obtaining site specific
 CC gene modification in cell lines.

XX Sequence 530 AA;

Query Match 77.1%; Score 2158.5; DB 21; Length 530;

Best Local Similarity 74.5%; Pred. No. 2.5e-202; Indels 1; Gaps 1;

Matches 395; Conservative 60; Mismatches 74; Indels 1; Gaps 1;

QY 1 MSKMTSALLLIQSCYSSGCGKVLVWPTFESHMNNKITLDELVORGHEVTVLASSA 60
 DB 1 MSKMTSALLLIQSCYSSGCGKVLVWPTFESHMNNKITLDELVORGHEVTVLASSA 60
 QY 61 SISFDNPSPTLKEVYVVSILTKTEFEDIIKQLYKRA-ELPKDTFMSYSQVOEIMWTF 119
 DB 61 STLVAASKSAIKLEVPTSLTKNDLSDLSLKILDRMIVGVSKTFMSYSQLOELCWEY 120
 QY 120 NDILRKFCKDIVSNKKLQESRFDVVLADAVFPPEGLAEILKIPFVYSLSRSPGYA 179
 DB 121 YDYSNKLCKDAVLNKKLMKLOESKFVYILADALNPGCELLAEILNPLLSLRFSGYT 180
 QY 180 IEKHSGLLPSPSYVPMVMSLSDQMTFIERVKNNIYLVYFEFQIFDMKKMDQFYSEV 239
 DB 181 FEKNGGGFLPPPSYVPMVMSLSDQMTFIERVKNNIYLVYFEFQIFDMKKMDQFYSEV 240
 QY 240 LGRPTTLESTMAKADIVLIRNYWDFOPPHLLPNVEVVGILHCKPAKPLPEMEEFVQS 299
 DB 241 LGRPTTLESTMAKADIVLIRNYWDFOPPHLLPNVEVVGILHCKPAKPLPEMEEFVQS 300
 QY 300 GENGIVVFSLSGMSVNTSEERANVIASALAKIPQKVLNRPFGNKPDTLGLNTRLYKMIPO 359
 DB 301 GENGIVVFSLSGMSVNTSEERANVIASALAKIPQKVLNRPFGNKPDTLGLNTRLYKMIPO 360
 QY 360 NDILGHPKTKAFITGGNGIYEALYHGVPMVGVPIFGDQLDNIAMKAKGAVEINFT 419
 DB 361 NDILGHPKTKAFITGGNGIYEALYHGVPMVGVPIFGDQLDNIAMKAKGAALSVDIRT 420
 QY 420 MTSDDLRLARTVITDSSYKENAMRLSRIHHDQVPKPLDRAVFIEMVRRHKGAKHLRSA 479
 DB 421 MSSRDLNLNALKSVINDPVYKENVMKLSRIHHDQPMKPLDRAVFIEMVRRHKGAKHLRSA 480
 QY 480 AHDLTWFOHYSIDVIGFLTCVATAIFLTKCFPLFSCQENKTKIEKRE 529
 DB 481 AHNLTWIOHSLDVIAFLACVATVITITKFCFLCFRKLAKTKGKKRD 530

RESULT 12

ABJ19806 standard; Protein; 530 AA.

XX ABJ19806;

DT 10-APR-2003 (first entry)

DB Androgen-independent prostate cancer-related protein - SEQ ID No 12.

XX Androgen-independent cancer; androgen ablation therapy; prostate cancer;

KW androgen-dependent prostate cancer; prostate cancer.

XX unidentified.

XX WO200298358-A2.

PD 12-DEC-2002.

PE 04-JUN-2002; 2002WO-US17594.

XX 04-JUN-2001; 2001US-295917P.

PR 13-NOV-2001; 2001US-350666P.

PR 29-MAR-2002; 2002US-368689P.

PR 12-APR-2002; 2002US-372246P.
 PR 31-MAY-2002; 2002US-0160233.
 XX (BOSB-) EOS BIOTECHNOLOGY INC.
 PA Afar DEH, Agus D, Mack DH;

XX WPI; 2003-148602/14.

XX Detecting an androgen-independent prostate cancer cell in a sample or

PT diagnosing androgen-dependent prostate cancer, by determining the

PT presence or absence of genes whose expressions are up- or

PT down-regulated -

XX Claim 1; Page 188; 210pp; English.

CC The invention comprises a method for detecting an androgen-independent
 CC cancer cell in a sample from a patient who has undergone androgen
 CC ablation therapy. The method involves determining the presence or absence
 CC of nucleic acids that are either up-regulated or down-regulated in
 CC prostate cancer. The method is useful for detecting an androgen-
 CC independent prostate cancer cell in a sample from a patient who has
 CC undergone androgen ablation therapy. The method is particularly useful
 CC for diagnosing androgen-dependent prostate cancer, prostate cancer
 CC undergoing androgen withdrawal, or androgen-independent prostate cancer.
 CC The present amino acid sequence represents a protein which is encoded by
 CC a gene that is either up-regulated or down-regulated in prostate cancer.

XX Sequence 530 AA;

Query Match 77.1%; Score 2158.5; DB 24; Length 530;

Best Local Similarity 74.5%; Pred. No. 2.5e-202; Indels 1; Gaps 1;

Matches 395; Conservative 60; Mismatches 74; Indels 1; Gaps 1;

QY 1 MSKMTSALLLIQSCYSSGCGKVLVWPTFESHMNNKITLDELVORGHEVTVLASSA 60
 DB 1 MSKMTSALLLIQSCYSSGCGKVLVWPTFESHMNNKITLDELVORGHEVTVLASSA 60
 QY 61 SISFDNPSPTLKEVYVVSILTKTEFEDIIKQLYKRA-ELPKDTFMSYSQVOEIMWTF 119
 DB 61 STLVAASKSAIKLEVPTSLTKNDLSDLSLKILDRMIVGVSKTFMSYSQLOELCWEY 120
 QY 120 NDILRKFCKDIVSNKKLQESRFDVVLADAVFPPEGLAEILKIPFVYSLSRSPGYA 179
 DB 121 YDYSNKLCKDAVLNKKLMKLOESKFVYILADALNPGCELLAEILNPLLSLRFSGYT 180
 QY 180 IEKHSGLLPSPSYVPMVMSLSDQMTFIERVKNNIYLVYFEFQIFDMKKMDQFYSEV 239
 DB 181 FEKNGGGFLPPPSYVPMVMSLSDQMTFIERVKNNIYLVYFEFQIFDMKKMDQFYSEV 240
 QY 240 LGRPTTLESTMAKADIVLIRNYWDFOPPHLLPNVEVVGILHCKPAKPLPEMEEFVQS 299
 DB 241 LGRPTTLESTMAKADIVLIRNYWDFOPPHLLPNVEVVGILHCKPAKPLPEMEEFVQS 300
 QY 300 GENGIVVFSLSGMSVNTSEERANVIASALAKIPQKVLNRPFGNKPDTLGLNTRLYKMIPO 359
 DB 301 GENGIVVFSLSGMSVNTSEERANVIASALAKIPQKVLNRPFGNKPDTLGLNTRLYKMIPO 360
 QY 360 NDILGHPKTKAFITGGNGIYEALYHGVPMVGVPIFGDQLDNIAMKAKGAVEINFT 419
 DB 361 NDILGHPKTKAFITGGNGIYEALYHGVPMVGVPIFGDQLDNIAMKAKGAALSVDIRT 420
 QY 420 MTSDDLRLARTVITDSSYKENAMRLSRIHHDQVPKPLDRAVFIEMVRRHKGAKHLRSA 479
 DB 421 MSSRDLNLNALKSVINDPVYKENVMKLSRIHHDQPMKPLDRAVFIEMVRRHKGAKHLRSA 480
 QY 480 AHDLTWFOHYSIDVIGFLTCVATAIFLTKCFPLFSCQENKTKIEKRE 529
 DB 481 AHNLTWIOHSLDVIAFLACVATVITITKFCFLCFRKLAKTKGKKRD 530

RESULT 13

AAU29284

ID AAU29284 standard; Protein; 527 AA.
XX
AC AAU29284;
XX
DT 18-DEC-2001 (first entry)
XX
XX Human PRO polypeptide sequence #261.
XX
KW PRO polypeptide; mammal; tumour; cancer; human; cattle; horse; sheep;
KW dog; cat; pig; goat; rabbit; tumour necrosis factor alpha; TNF-alpha;
KW blood; chondrocyte cell; cell proliferation; cell differentiation; colon;
KW adrenal; lung; breast; prostate; rectum; cervix; liver; genetic disorder.
XX
OS Homo sapiens.
XX
PN WO200168848-A2.
XX
PD 20-SEP-2001.
XX
PF 28-FEB-2001; 2001WO-US06520.
XX
PR 01-MAR-2000; 2000WO-US05601.
PR 02-MAR-2000; 2000WO-US05841.
PR 03-MAR-2000; 2000US-187202P.
PR 06-MAR-2000; 2000US-186968P.
PR 14-MAR-2000; 2000US-189320P.
PR 14-MAR-2000; 2000US-189328P.
PR 15-MAR-2000; 2000WO-US06884.
PR 21-MAR-2000; 2000US-190828P.
PR 21-MAR-2000; 2000US-191077P.
PR 21-MAR-2000; 2000US-191048P.
PR 21-MAR-2000; 2000US-19114P.
PR 28-MAR-2000; 2000US-192655P.
PR 29-MAR-2000; 2000US-193032P.
PR 29-MAR-2000; 2000US-193033P.
PR 30-MAR-2000; 2000WO-US08439.
PR 04-APR-2000; 2000US-194449P.
PR 04-APR-2000; 2000US-194647P.
PR 11-APR-2000; 2000US-195975P.
PR 11-APR-2000; 2000US-196000P.
PR 11-APR-2000; 2000US-196187P.
PR 11-APR-2000; 2000US-196690P.
PR 11-APR-2000; 2000US-196820P.
PR 18-APR-2000; 2000US-198121P.
PR 18-APR-2000; 2000US-198585P.
PR 25-APR-2000; 2000US-199397P.
PR 25-APR-2000; 2000US-199509P.
PR 25-APR-2000; 2000US-199549P.
PR 03-MAY-2000; 2000US-201516P.
PR 17-MAY-2000; 2000WO-US13705.
PR 22-MAY-2000; 2000WO-US14042.
PR 30-MAY-2000; 2000WO-US14941.
PR 02-JUN-2000; 2000WO-US15264.
PR 05-JUN-2000; 2000US-209832P.
PR 28-JUL-2000; 2000WO-US20710.
PR 22-AUG-2000; 2000US-064484P.
PR 24-AUG-2000; 2000WO-US23328.
PR 08-NOV-2000; 2000WO-US30952.
PR 01-DEC-2000; 2000WO-US32678.
PR 20-DEC-2000; 2000WO-US34956.
XX
PA (GENTH) GENENTECH INC.
XX
PI Baker KP, Chen J, Desnoyers L, Goddard A, Godowski PJ, Gurney AL;
PI Pan J, Smith V, Watanabe CK, Wood WI, Zhang Z;
XX
XX MPI; 2001-602746/68.
XX
XX N-PSDB; AAS46185.
XX
XX Novel nucleic acids encoding PRO polypeptides, used to diagnose the
XX presence of tumours, such as prostate and breast tumours, in mammals and
XX to screen for modulators of the compounds -
XX

PS Claim 11; Fig 522; 774bp; English.
XX
XX Sequences AAU29284-AAU29328 represent PRO polypeptides of the invention.
CC The PRO polypeptides and their associated nucleic acids can be used to
CC detect the presence of a tumour in a mammal by comparing the level of
CC expression of a PRO polypeptide in a test sample of cells from the animal
CC and a control sample of normal cells, whereby a higher level of
CC expression in the test sample indicates the presence of a tumour in the
CC mammal. Mammals include dogs, cats, cattle, horses, sheep, pigs, goats
CC and rabbits but are preferably human. The polypeptides can be used to
CC stimulate tumour necrosis factor (TNF) alpha release from human blood,
CC when contacted with it. A specific polypeptide can be used to stimulate
CC the proliferation or differentiation of chondrocyte cells. The PRO
CC proteins can be used to determine the presence of tumours and also
CC susceptibility to tumour development, particularly adrenal, lung, colon,
CC breast, prostate, rectal, cervical, or liver tumours, in mammalian
CC subjects. The oligonucleotide probes specific for the PRO nucleic acids
CC can be used for genetic analysis of individuals with genetic disorders.
XX
XX
SQ Sequence 527 AA;
Query Match 68.8%; Score 1924.5; DB 22; Length 527;
Best Local Similarity 69.9%; Pred. No. 2.2e-179;
Matches 369; Conservative 47; Mismatches 109; Indels 3; Gaps 2;
QY 2 SMKWTSAIIILITLSCEYSSGSCGKLVWPTFESHMNIKITLDELVOGHEVTYLAASAS 61
DB 3 SDRKALVFLILQLFC-VGCGRCGKLVWPCDMSHMLAVKVLLEELVIRGHEVTVLTSHKP 61
QY 62 ISFDPSNPSTLKEFVYVSLTKTEFEDIIKQLVKRWAELEPKDTWSPVSOQELMTWTFND 121
DB 62 SLIDYRKPSALKFEVYVHMPDRTEENEFVDLALN-VLPBLSTWQSVIKLNDFFVEIRG 119
QY 122 ILRKPCDVIENKRLMKLOESRFDVVLADAVPEFGELELTKIPFYYSRFGYALE 181
DB 120 TLKWCSEFIYNQTLMKLOETNYDVMILDIVIPCGLMELLAVFVLTLRISVGNGME 179
QY 182 KHSGGLFPPSPYVNVNSELSDQMTFIERVNMIVLVYFEFWFOIPDMKKWDQFSEVTLG 241
DB 180 RSCGKLPAPLSYVVPMTGLDRMTFLERVNSMLSVLFHWMIDYDHFWEEFYSKALG 239
QY 242 RPTLSETMAKADILIRNWDPOFPHPLDNVEFVGGHCKPAKPLPKEMEETVQSSGE 301
DB 240 RPTLCETVGAKEIWLIRTYWDFEFPPOYPONFEFVGLHCKPAKALPKEMENFVQSSGE 299
QY 302 NGVVVFSLSGWSVNTSEBAAVIVASALAKIPOKLWMPDGKRPDLGINTFLYKPIPN 361
DB 300 DGIIVFSLGSLFQVNTBKKANIIASALAQIPQKVLWKYKGGKBSILGANTRLYDIPND 359
QY 362 LLGHPKTKAFTTHGAMNGIYEAIVHGVPMVGPPIFGDQLDNIAMKKAAGAAVEINFKMT 421
DB 360 LLGHPKTKAFTTHGAMNGIYEAIVHGVPMVGPPIFGDQLDNIAMKKAAGAAVEINFKMT 419
QY 422 SEDILRALRIVITDSSYKENMRLSRHHDPVYKPLDRAVWIEFWMRHKAHLRSAAH 481
DB 420 SEDILRALRIVITDSSYKENMRLSRHHDPVYKPLDRAVWIEFWMRHKAHLRSAAH 479
QY 482 DLTWFOHYSIDVIGFLTCTVATLFLFKCGLFSCQENKTRKLEKE 529
DB 480 DLTWFOHYSIDVIGFLTCTVATLFLFKCGLFSCQENKTRKLEKE 527
RESULT 14
AAU7927
ID AAU7927 standard; Protein; 527 AA.
XX
XX AAU7927;
XX
XX 18-JUN-2002 (first entry)
XX
XX Human drug-metabolising enzyme.
XX
XX Human drug-metabolising enzyme; chromosome 4; therapeutic agent;
XX
XX

DR WPI: 2002-066363/09.
 DR N-PSDB: AAD24666.
 XX
 PT Novel isolated human drug metabolizing enzymes referred as DME 1-10,
 PT useful for diagnosing, treating, or preventing disorders associated
 PT with aberrant expression of DME such as allergy, anemia, asthma,
 PT infertility -

PS Claim 1a; Page 126-127; 143pp; English.

XX
 CC The invention relates to human drug metabolizing enzymes referred as
 CC DME and nucleic acid molecules encoding such enzymes. Polynucleotides
 CC of the invention are useful for assessing toxicity of test compounds
 CC and in gene therapy. Sequences of the invention are useful in the
 CC diagnosis, prevention and treatment of autoimmune/inflammatory
 CC disorders such as acquired immune deficiency syndrome (AIDS), adult
 CC respiratory distress syndrome, allergies, anaemia, atherosclerosis,
 CC asthma, autoimmune haemolytic anaemia, contact dermatitis, Crohn's
 CC disease, glomerulonephritis, Goodpasture's syndrome, gout, Grave's
 CC disease, Hashimoto's thyroiditis, irritable bowel syndrome, multiple
 CC sclerosis, osteoarthritis, osteoporosis, psoriasis, systemic lupus
 CC erythematosus, rheumatoid arthritis, scleroderma, ulcerative colitis,
 CC uveitis, viral, bacterial, fungal, parasitic, protozoal, helminthic
 CC infections; cell proliferative disorders such as actinic keratosis,
 CC arteriosclerosis, atherosclerosis, Duchenne and Becker dystrophy,
 CC hepatitis, Cushing's syndrome, cancers, myelodysplastic syndrome,
 CC epilepsy; endocrine disorders such as disorders of the hypothalamus
 CC and pituitary resulting from lesions such as primary brain tumours,
 CC adenomas, infarction associated with pregnancy, aneurysms, vascular
 CC malformations; eye disorders such as conjunctivitis, iritis, retinitis,
 CC glaucoma, pigmentosa; metabolic disorders such as Addison's disease,
 CC cystic fibrosis, diabetes, goitre, glycogen storage diseases,
 CC hypercholesterolaemia, hyperthyroidism, hypoglycaemia, lipid myopathies,
 CC Menkes syndrome, mannosidosis, obesity; gastrointestinal disorders such
 CC as dysphagia, gastric carcinoma, anorexia, nausea, gastroenteritis,
 CC hyperbilirubinaemia, emesis, cirrhosis, diarrhoea, jaundice, Reye's
 CC syndrome, peliosis hepatitis, hepatic vein thrombosis and developmental
 CC disorders. The present sequence is human DME-1 protein.

XX Sequence 527 AA;

Query Match 68.8%; Score 1924.5; DB 23; Length 527;
 Best Local Similarity 69.9%; Pred. No. 2.2e-179;
 Matches 369; Conservative 47; Mismatches 109; Indels 3; Gaps 2;

QY 2 SMKWTSAALLIQLSCTYFSSGCGKVLVPTFESHMNIKTIDELVORGHEVTLASAS 61
 DB 3 SDKSALVFLQLFC-VCGGFCGKVLVPCDMSHMLNVKILIELIVRGHEVTLTHSKP 61
 QY 62 ISFDNPSSTLKEEYVPVSLTKEFEEDIIKQVYKMAELPKDTFWSYFSGVQEIIMTFND 121
 DB 62 SLIDTRKPSALKFEYVHWPQDRTENEIFVDLALN-VLPGLSHWOSVIKLNDFFVEIRG 119
 QY 122 ILKRFCKDIVSNKMLKQLQESRFDVLADAVFPFGLAELDLKIPVYSLSRFSPOVAIE 181
 DB 120 TLKMWCESTFYQTLMKLQETNYVMDLPYIPCGDLMAELAVPVLTLRISVGANME 179
 QY 182 KHSGLLEPPSYVYVPMSELSDQMTFIERVKMIVLYFEFWQIFDMKKMDQFYSEVLG 241
 DB 180 RSCGKLPAPLSYVPMVMTGLTRMTFLERVKMSMLSVLFHWIODYDYHFWEEFYSKALG 239
 QY 242 RPTTSETMAKADIMLRNYMDFOPRPHLLPNVEFVGGLHCKPAKPLPKMEEFVQSSGE 301
 DB 240 RPTTCEVVGKAEIMLRNYMDFOPRPHLLPNVEFVGGLHCKPAKPLPKMEEFVQSSGE 299
 QY 302 NGVVFSLGSMTSNTSEBRANVIASLAKIPKVLNRPDGNKPDITGLNTRLYKWIPOND 361
 DB 300 DGIIVFSLGSLFQNTVEKANIISALAOIPKVLNRYKGRKPSLIGANTRLYDWIPOND 359
 QY 362 LLGHPKTAFTTHGNGGIYEAIIYGVPMVGPVIFGDQIDNIAMKAKGAIVEINFKTWT 421
 DB 360 LLGHPKTAFTTHGNGGIYEAIIYGVPMVGPVIFGDQIDNIAMKAKGAIVEINFKTWT 419

QY 422 SEDLLRALRTVITDSSYKENAMRLSRIHDDQPVKPLDRAVFWIEFWNRHKGALRSAAH 481
 DB 420 SEDLLRALRTVITDSSYKENAMRLSRIHDDQPVKPLDRAVFWIEFWNRHKGALRSAAH 479
 QY 482 DLTWFOHYSIDVIGFLITCVATAIPLFTKCFLLSCQKFNTRIKERE 529
 DB 480 DLTWFOHYSIDVIGFLITCVATAIPLFTKCFLLSCQKFNTRIKERE 527

Search completed: December 5, 2003, 09:48:21
 Job time : 49 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 5, 2003, 09:46:35 ; Search time 21 Seconds

(without alignments)
2422.535 Million cell updates/sec

Title: US-09-980-729B-5

Perfect score: 2798
Sequence: 1 MSKMTSALLLIQUSCFYSS.....KCFIFSCQKFKTRIKRE 529

Scoring table: BIOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :
1: PIR_76:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|--------|-------------|--------|----------|--------------------|
| 1 | 2590.5 | 92.6 | 528 | 2 JN0619 | glucuronosyltransf |
| 2 | 2330 | 83.3 | 529 | 2 A35366 | glucuronosyltransf |
| 3 | 2287 | 81.7 | 529 | 2 J50200 | orphan UDP-glucuro |
| 4 | 2281.5 | 81.5 | 528 | 2 JN0620 | UDP-glucuronosyltr |
| 5 | 2158.5 | 77.1 | 530 | 2 A48633 | glucuronosyltransf |
| 6 | 2086.5 | 74.6 | 523 | 2 S11309 | glucuronosyltransf |
| 7 | 1991.5 | 71.2 | 520 | 2 C47113 | glucuronosyltransf |
| 8 | 1975.5 | 70.6 | 529 | 2 A42233 | glucuronosyltransf |
| 9 | 1938 | 69.3 | 531 | 2 B47113 | glucuronosyltransf |
| 10 | 1930.5 | 69.0 | 530 | 2 S68200 | glucuronosyltransf |
| 11 | 1904.5 | 68.1 | 530 | 2 S00163 | glucuronosyltransf |
| 12 | 1892.5 | 67.6 | 530 | 2 S07390 | glucuronosyltransf |
| 13 | 1853.5 | 66.2 | 530 | 2 A40467 | glucuronosyltransf |
| 14 | 1840.5 | 65.8 | 530 | 2 A36276 | glucuronosyltransf |
| 15 | 1760.5 | 62.9 | 527 | 2 S15089 | glucuronosyltransf |
| 16 | 1196.5 | 42.8 | 531 | 2 A35343 | glucuronosyltransf |
| 17 | 1137.5 | 40.7 | 533 | 2 A39092 | glucuronosyltransf |
| 18 | 1135.5 | 40.6 | 530 | 2 S15764 | glucuronosyltransf |
| 19 | 1100.5 | 39.3 | 530 | 2 J55656 | glucuronosyltransf |
| 20 | 1082.5 | 38.7 | 530 | 2 S17512 | glucuronosyltransf |
| 21 | 1064 | 38.0 | 531 | 2 A31360 | glucuronosyltransf |
| 22 | 1057 | 37.8 | 531 | 2 A55788 | glucuronosyltransf |
| 23 | 1045.5 | 37.4 | 529 | 2 A24600 | glucuronosyltransf |
| 24 | 1045 | 37.3 | 531 | 2 B55788 | glucuronosyltransf |
| 25 | 812 | 29.0 | 541 | 2 J54423 | 2-hydroxyacylsphn |
| 26 | 810.5 | 29.0 | 541 | 2 A48801 | 2-hydroxyacylsphn |
| 27 | 765 | 27.3 | 174 | 2 A47113 | UDP-glucuronosyltr |
| 28 | 554 | 19.8 | 508 | 2 T03910 | UDP-glucuronosyltr |
| 29 | 540.5 | 19.3 | 531 | 2 T23835 | hypothetical prote |

| | | | | | |
|----|-------|------|-----|----------|--------------------|
| 30 | 490 | 17.5 | 534 | 2 T19944 | hypothetical prote |
| 31 | 489 | 17.5 | 485 | 2 T13694 | glucuronosyltransf |
| 32 | 478.5 | 17.1 | 537 | 2 T21823 | hypothetical prote |
| 33 | 453 | 16.2 | 573 | 2 T27578 | hypothetical prote |
| 34 | 442 | 15.8 | 534 | 2 T19951 | hypothetical prote |
| 35 | 439 | 15.7 | 586 | 2 T19075 | hypothetical prote |
| 36 | 433.5 | 15.5 | 502 | 2 T25263 | hypothetical prote |
| 37 | 428 | 15.3 | 949 | 2 T18591 | hypothetical prote |
| 38 | 427 | 15.3 | 530 | 2 T19365 | hypothetical prote |
| 39 | 423 | 15.1 | 745 | 2 T23893 | hypothetical prote |
| 40 | 419 | 15.0 | 534 | 2 T34455 | hypothetical prote |
| 41 | 410.5 | 14.7 | 515 | 2 T32217 | hypothetical prote |
| 42 | 409 | 14.6 | 579 | 2 H88632 | protein F56B3.7 [i |
| 43 | 404.5 | 14.5 | 661 | 2 T32518 | hypothetical prote |
| 44 | 403 | 14.4 | 515 | 2 S52453 | ecdysteroid UDP-gl |
| 45 | 400 | 14.3 | 580 | 2 T28725 | hypothetical prote |

ALIGNMENTS

RESULT 1

glucuronosyltransferase (EC 2.4.1.17) 2B-4 precursor - human
JN0619
N/Alternate names: UDP-glucuronosyltransferase 2B-11
C/Species: Homo sapiens (man)
C/Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 02-Jun-2000
C/Accession: JN0619; A27878
R/Jin, C.U.; Miners, J.O.; Lillywhite, K.J.; Mackenzie, P.I.
Biochem. Biophys. Res. Commun. 194, 496-503, 1993
A/Title: Cloning of a human liver microsomal UDP-glucuronosyltransferase cDNA.
A/Reference number: JN0619; MUID:93326164; PMID:8333863
A/Accession: JN0619
A/Molecule type: mRNA
A/Residues: 1-528 <JIN>
A/Cross-references: GB:AF081793; NID:G3426331
A/Experimental source: liver
R/Jackson, M.R.; McCarthy, L.R.; Harding, D.; Wilson, S.; Coughtrie, M.W.H.; Burchell, B
Biochem. J. 242, 581-588, 1987
A/Title: Cloning of a human liver microsomal UDP-glucuronosyltransferase cDNA.
A/Reference number: A27878; MUID:87241362; PMID:3109396
A/Accession: A27878
A/Molecule type: mRNA
A/Residues: 1-108, 'F', 110-170, 'RP', 173-381, 'K', 383-384, 'SPR', 388-395, 'F', 397-528 <JAC>
A/Cross-references: GB:Y00317; NID:G37588; PIDN:CAA68415.1; PID:G37589
C/Genetics:
A/Gene: GDB:UGT2B4; UGT2B11
A/Cross-references: GDB:5891331; OMIM:600067
A/Map position: 4q13-4q13
C/Keywords: glucuronosyltransferase
C/Keywords: glycoprotein; glycosyltransferase; hexosyltransferase; transmembrane protein
F/1-23/Domain: signal sequence #status predicted <SIG>
F/24-528/Product: glucuronosyltransferase 2B-11 #status predicted <MET>
F/492-509/Domain: transmembrane #status predicted <TM>
F/315/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 92.6%; Score 2590.5; DB 2; Length 528;
Best Local Similarity 92.6%; Pred. No. 7.2e-192;
Matches 490; Conservative 14; Mismatches 24; Indels 1; Gaps 1;

QY 1 MSKMTSALLLIQUSCFYSSGCGKVLWPTFSHMNITITIDELVQRGHEVTVLASSA 60
DB 1 MSKMTSALLLIQUSCFYSSGCGKVLWPTFSHMNITITIDELVQRGHEVTVLASSA 60
QY 61 SIFSDPNSPTLKEFYVPSVLITFEEDIIKQLVKRWALPKDTFWSYFSQVOEIMWTFN 120
DB 61 SIFSDPNSPTLKEFYVPSVLITFEEDIIKQLVKRWALPKDTFWSYFSQVOEIMWTFN 120
QY 121 DLRKPKQIVSKMKLQESRFPVVLADAVPPGELLAELKLPVYSIAFSPGYAI 180
DB 121 DLRKPKQIVSKMKLQESRFPVVLADAVPPGELLAELKLPVYSIAFSPGYAI 180
QY 121 DLRKPKQIVSKMKLQESRFPVVLADAVPPGELLAELKLPVYSIAFSPGYAI 180
DB 121 DLRKPKQIVSKMKLQESRFPVVLADAVPPGELLAELKLPVYSIAFSPGYAI 180
QY 181 EKHSGLFPSPVVPVMSLSDQMTFIERKMNITVLYEFWFQFLDMKMDQFSEVL 240

```

Db 181 EKHSGLFPSPSYVPMVSELSDDMTFIERVYKMIYVLYFFPFMQIPFMKMDQFYSVL 240
Qy 241 GRPTLSTMAKADIWLRNWDFOFPHPLPNVFEVGLHCKPAKPLPKMEEFVSSG 300
Db 241 GRPTLSTMAKADIWLRNWDFOFPHPLPNVFEVGLHCKPAKPLPKMEEFVSSG 300
Qy 301 ENGIVVPSLGSVNSVTSERANVIASALAKIPQKVLMPDGNKPTDGLNRLYKWIPO 360
Db 301 ENGIVVPSLGSVNSVTSERANVIASALAKIPQKVLMPDGNKPTDGLNRLYKWIPO 360
Qy 361 DLGHPKTAFTTHGNGNGIYEALYHGVPMGVPIFGDQDNIAMKAKGAIVEINFTM 420
Db 361 DLGHPKTAFTTHGNGNGIYEALYHGVPMGVPIFGDQDNIAMKAKGAIVEINFTM 420
Qy 421 TSBDLRALRTVITDSSYKENAMRLSRHHDPVKPLDRAVFWIEFVWRHKGAKHLRVA 480
Db 421 TSBDLRALRTVITDSSYKENAMRLSRHHDPVKPLDRAVFWIEFVWRHKGAKHLRVA 480
Qy 481 HDLTFPHYSIDVIGFLTCAVTAIFLTKCFLFCCKFKNTRKIEKRE 529
Db 481 HDLTFPHYSIDVIGFLTCAVTAIFLTKCFLFCCKFKNTRKIEKRE 529

```

RESULT 2

```

A35366
glucuronosyltransferase (EC 2.4.1.17) UDPGth-2 precursor - human
C/Species: Homo sapiens (man)
C/Date: 17-Aug-1990 #sequence_revision 17-Aug-1990 #text_change 29-Sep-1999
C/Accession: A35366
R/Ritter, J.K.; Sheen, Y.Y.; Owens, I.S.
J. Biol. Chem. 265: 7900-7906, 1990
A/Title: Cloning and expression of human liver UDP-glucuronosyltransferase in COS-1 cell
A/Reference number: A35366; MUID:90243659; PMID:2159463
A/Accession: A35366
A/Status: preliminary
A/Molecule type: mRNA
A/Residues: 1-529 <RIT>
A/Cross-references: GB:J05428; NID:G340079; PIDN:AAA36793.1; PID:G340080
C/Genetics:
A/Gene: GDB:UGT2B7; UGT2B9
A/Cross-references: GDB:5892203; OMIM:600218
A/Map position: 4q13-4q13
C/Superfamily: glucuronosyltransferase
C/Keywords: glycosyltransferase; hexosyltransferase; transmembrane protein

```

Query Match 83.3%; Score 2330; DB 2; Length 529;

Best Local Similarity 81.1%; Pred. No. 9, 1e-172;

Matches 429; Conservative 42; Mismatches 58; Indels 0; Gaps 0;

```

Qy 1 MSKMTSALLLIQLSCYSSGCGKVLVWPFEPFHMNIKITLDLVORGHVTVLASSA 60
Db 1 MSKMTSALLLIQLSCYSSGCGKVLVWPFEPFHMNIKITLDLVORGHVTVLASSA 60
Qy 61 SISFPNPSPTLKFEVYVPSLTKEFEDIKQLVKMAELPKDTFMSYFSQVQEIIMTFN 120
Db 61 SISFPNPSPTLKFEVYVPSLTKEFEDIKQLVKMAELPKDTFMSYFSQVQEIIMTFN 120
Qy 121 DILRPFCKDIYSNKKLKKQESRPDVYLAADVFPFGBELABELLKIPVYSLSRSPGYAI 180
Db 121 DILRPFCKDIYSNKKLKKQESRPDVYLAADVFPFGBELABELLKIPVYSLSRSPGYAI 180
Qy 181 EKHSGLFPSPSYVPMVSELSDDMTFIERVYKMIYVLYFFPFMQIPFMKMDQFYSVL 240
Db 181 EKHSGLFPSPSYVPMVSELSDDMTFIERVYKMIYVLYFFPFMQIPFMKMDQFYSVL 240
Qy 241 GRPTLSTMAKADIWLRNWDFOFPHPLPNVFEVGLHCKPAKPLPKMEEFVSSG 300
Db 241 GRPTLSTMAKADIWLRNWDFOFPHPLPNVFEVGLHCKPAKPLPKMEEFVSSG 300
Qy 301 ENGIVVPSLGSVNSVTSERANVIASALAKIPQKVLMPDGNKPTDGLNRLYKWIPO 360
Db 301 ENGIVVPSLGSVNSVTSERANVIASALAKIPQKVLMPDGNKPTDGLNRLYKWIPO 360

```

```

Qy 361 DLGHPKTAFTTHGNGNGIYEALYHGVPMGVPIFGDQDNIAMKAKGAIVEINFTM 420
Db 361 DLGHPKTAFTTHGNGNGIYEALYHGVPMGVPIFGDQDNIAMKAKGAIVEINFTM 420
Qy 421 TSBDLRALRTVITDSSYKENAMRLSRHHDPVKPLDRAVFWIEFVWRHKGAKHLRVA 480
Db 421 TSBDLRALRTVITDSSYKENAMRLSRHHDPVKPLDRAVFWIEFVWRHKGAKHLRVA 480
Qy 481 HDLTFPHYSIDVIGFLTCAVTAIFLTKCFLFCCKFKNTRKIEKRE 529
Db 481 HDLTFPHYSIDVIGFLTCAVTAIFLTKCFLFCCKFKNTRKIEKRE 529

```

RESULT 3

```

JE0200
orphan UDP-glucuronosyltransferase (EC 2.4.-.-) - human
N/Alternate names: UGT2B11
C/Species: Homo sapiens (man)
C/Date: 21-Aug-1998 #sequence_revision 21-Aug-1998 #text_change 17-Mar-1999
C/Accession: JE0200
R/Beaulieu, M.; Levesque, E.W.; Hum, D.; Belanger, A.
Biochem. Biophys. Res. Commun. 248: 44-50, 1998
A/Title: Isolation and characterization of a human orphan UDP-glucuronosyltransferase, U
A/Reference number: JE0200; MUID:98340847; PMID:9675083
A/Accession: JE0200
A/Molecule type: mRNA
A/Residues: 1-529 <BRA>
A/Cross-references: GB:AF016492
C/Comment: This enzyme catalyzes the transfer of glucuronic acid to a wide range of exog
C/Superfamily: glucuronosyltransferase
C/Keywords: glycoprotein; glycosyltransferase
F/493-509/domain: membrane-anchoring #status predicted <MC>
F/315/Binding site: carbohydrate (Asn) (covalent) #status predicted

```

Query Match 81.7%; Score 2287; DB 2; Length 529;

Best Local Similarity 80.0%; Pred. No. 1, 9e-168;

Matches 423; Conservative 43; Mismatches 63; Indels 0; Gaps 0;

```

Qy 1 MSKMTSALLLIQLSCYSSGCGKVLVWPFEPFHMNIKITLDLVORGHVTVLASSA 60
Db 1 MTLKMTSVLLIHLSCYSSGCGKVLVWMAEYSHMMNKITLDELVORGHVTVLASSA 60
Qy 61 SISFPNPSPTLKFEVYVPSLTKEFEDIKQLVKMAELPKDTFMSYFSQVQEIIMTFN 120
Db 61 SILDPNDASTLKKEVYVPSLTKEFEDIKQVKKSDIKDSFMYLFSQEIIMELV 120
Qy 121 DILRPFCKDIYSNKKLKKQESRPDVYLAADVFPFGBELABELLKIPVYSLSRSPGYAI 180
Db 121 DIFNPFCKDVSNKKKKQESRPDIYFADAVFPFGBELABELLKIPVYSLSRSPGYAI 180
Qy 181 EKHSGLFPSPSYVPMVSELSDDMTFIERVYKMIYVLYFFPFMQIPFMKMDQFYSVL 240
Db 181 ERHSGGLFPSPSYVPMVSELSDDMTFIERVYKMIYVLYFFPFMQIPFMKMDQFYSVL 240
Qy 241 GRPTLSTMAKADIWLRNWDFOFPHPLPNVFEVGLHCKPAKPLPKMEEFVSSG 300
Db 241 GRPTLSTMAKADIWLRNWDFOFPHPLPNVFEVGLHCKPAKPLPKMEEFVSSG 300
Qy 301 ENGIVVPSLGSVNSVTSERANVIASALAKIPQKVLMPDGNKPTDGLNRLYKWIPO 360
Db 301 ENGIVVPSLGSVNSVTSERANVIASALAKIPQKVLMPDGNKPTDGLNRLYKWIPO 360
Qy 361 DLGHPKTAFTTHGNGNGIYEALYHGVPMGVPIFGDQDNIAMKAKGAIVEINFTM 420
Db 361 DLGHPKTAFTTHGNGNGIYEALYHGVPMGVPIFGDQDNIAMKAKGAIVEINFTM 420
Qy 421 TSBDLRALRTVITDSSYKENAMRLSRHHDPVKPLDRAVFWIEFVWRHKGAKHLRVA 480
Db 421 TSBDLRALRTVITDSSYKENAMRLSRHHDPVKPLDRAVFWIEFVWRHKGAKHLRVA 480
Qy 481 HDLTFPHYSIDVIGFLTCAVTAIFLTKCFLFCCKFKNTRKIEKRE 529
Db 481 HDLTFPHYSIDVIGFLTCAVTAIFLTKCFLFCCKFKNTRKIEKRE 529

```

Db 481 HDLTFWQYHSLDVIGFLACVATVFIITKCELCFCFMKPAKGGKKRD 529

RESULT 4

JN0620

UDP-glucuronosyltransferase (EC 2.4.1.-) 2B-10 precursor - human
C/Species: Homo sapiens (man)
C/Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 28-May-1999
C/Accession: JN0620
R:Jin, C.J.; Miners, J.O.; Lillywhite, K.J.; Mackenzie, P.I.
Biochem. Biophys. Res. Commun. 194, 496-503, 1993
A/Title: cDNA cloning and expression of two new members of the human liver UDP-glucuronosyltransferase family: JN0619, MUID:93326164, PMID:8333863
A/Reference number: JN0619, MUID:93326164, PMID:8333863
A/Accession: JN0620
A/Molecule type: mRNA
A/Residues: 1-528 <JIN>
A/Cross-references: GB:X63359, NID:G516149, PIDN:CAA44961.1, PID:G516150
A/Experimental source: liver
C/Superfamily: glucuronosyltransferase; hexosyltransferase; transmembrane protein
C/Keywords: glycoprotein; glycosyltransferase; hexosyltransferase; transmembrane protein
F:1-23/Domain: signal sequence #status predicted <SIG>
F:24-528/Product: UDP-glucuronosyltransferase 2B-10 #status predicted <MET>
F:491-508/Domain: transmembrane #status predicted <TM>
F:66,314,481/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 81.5%; Score 2281.5; DB 2; Length 528;

Best Local Similarity 80.5%; Pred. No. 5e-168; Mismatches 44; Indels 1; Gaps 1;

Matches 426; Conservative 44; Mismatches 56; Indels 1; Gaps 1;

Qy 1 MSMKTSALLLIQLSCYSSGCGKLVLPTEFSHMNIKITILDELVQGHVTLASSA 60
Db 1 MALKMTT-VLLIQSLFYSSGCGKLVMAEYSLMMNMKITILKELVQGHVTLASSA 59
Qy 61 SISFDNPSPTLKEVYVVSITKTEFEDIIKQLVKRAELPKDTFWSYFSQVQELMTWF 120
Db 60 SILDPNDSSLTLEEVPTSLTKTEFENIMQLVKRSLEIQDPTFWSYFSQVQELMTWF 119
Qy 121 DILKRCFCDIYVSNKKMLKQESRPDYVADAVFPFGELLAELKIPVYSLRSPGYAI 180
Db 120 DIINFCQDVSNKKMLKQESRPDYVADAVFPFGELLAELKIPVYSHSFSFGYSF 179
Qy 181 EKHSQGLLPPSPYVPMVSELSDOMTFIERVKNMIVLYFEFVFOIFDMKKMDQFYSEV 240
Db 180 ERHSGGFPPSPYVPMVSELSDOMTFIERVKNMIVLYFEFVFOIFDMKKMDQFYSEV 239
Qy 241 GRPTTLSETMAKADIWLIRNYWDFPRLPLPNVEFVGLHCKPAKPLPKMEEFVQSS 300
Db 240 GRPTTLSETMAKADIWLIRNYWDFPRLPLPNVEFVGLHCKPAKPLPKMEEFVQSS 299
Qy 301 ENGIVVSLGSMVNTSEERANVIALAKIPQVLMRFDGNKPDGLNTRLYKMIPO 360
Db 300 ENGIVVSLGSMVNTSEERANVIALAKIPQVLMRFDGNKPDGLNTRLYKMIPO 359
Qy 361 DILGHPKTAFTTHGGMGIYEALYHGVPMVGPVIFGDDLDNIAMKAKAAVEINFTM 420
Db 360 DILGHPKTAFTTHGGMGIYEALYHGVPMVGPVIFGDDLDNIAMKAKAAVEINFTM 419
Qy 421 TSEDLRALRTVITDSSYKENAMRLSRHHDPVYKPLDRAVFWIEFVMBHKGAKLRSA 480
Db 420 TSEDLRALRTVITDSSYKENAMRLSRHHDPVYKPLDRAVFWIEFVMBHKGAKLRSA 479
Qy 481 HDLTFWQYHSLDVIGFLACVATVFIITKCELCFCFMKPAKGGKKRD 529
Db 480 HDLTFWQYHSLDVIGFLACVATVFIITKCELCFCFMKPAKGGKKRD 528

RESULT 5

A48633

glucuronosyltransferase (EC 2.4.1.17) precursor - human
N/Alternate names: dihydrotestosterone/androstenediol UDP-glucuronosyltransferase isoform
C/Species: Homo sapiens (man)
C/Date: 07-Apr-1994 #sequence_revision 18-Nov-1994 #text_change 29-Sep-1999
C/Accession: A48633, I38559

R:Chen, F.; Ritter, J.K.; Wang, M.G.; McBride, O.W.; Lubet, R.A.; Owens, I.S.

Biochemistry 32, 10648-10657, 1993

A/Title: Characterization of a cloned human dihydrotestosterone/androstenediol UDP-glucuronosyltransferase

A/Reference number: A48633, MUID:94002056, PMID:8399210

A/Accession: A48633

A/Status: preliminary; not compared with conceptual translation

A/Molecule type: nucleic acid

A/Residues: 1-530 <CH>

A/Experimental source: liver

A/Note: sequence extracted from NCBI backbone (NCBI:P138786)

R:Green, M.D.; Oturu, E.M.; Tepily, T.R.

Drug Metab. Dispos. 22, 799-805, 1994

A/Title: Stable expression of a human liver UDP-glucuronosyltransferase (UGT2B15) with a

A/Reference number: I38559, MUID:95136867, PMID:7835232

A/Accession: I38559

A/Status: preliminary; translated from GB/EMBL/DBJ

A/Molecule type: mRNA

A/Residues: 1-530 <RES>

A/Cross-references: EMBL:U08854, NID:9475758, PIDN:AAC50077.1, PID:9475759

C/Genetics:

A/Gene: GDB:UGT2B15, UGT2B8

A/Cross-references: GDB:5892418, OMIM:600219

A/Map position: 4q13-4q13

C/Superfamily: glucuronosyltransferase

C/Keywords: glycosyltransferase; hexosyltransferase

Query Match 77.1%; Score 2158.5; DB 2; Length 530;

Best Local Similarity 74.5%; Pred. No. 1.6e-158; Mismatches 395; Conservative 60; Indels 1; Gaps 1;

Matches 395; Conservative 60; Mismatches 74; Indels 1; Gaps 1;

Qy 1 MSMKTSALLLIQLSCYSSGCGKLVLPTEFSHMNIKITILDELVQGHVTLASSA 60
Db 1 MSUKTIVPFLIQLSCYSSGCGKLVLPTEFSHMNIKITILDELVQGHVTLASSA 60
Qy 61 SISFDNPSPTLKEVYVVSITKTEFEDIIKQLVKRAELPKDTFWSYFSQVQELMTWF 119
Db 61 STLVNASKSAIKLEEVPTSLTKTEFEDIIKQLVKRAELPKDTFWSYFSQVQELMTWF 120
Qy 121 DILKRCFCDIYVSNKKMLKQESRPDYVADAVFPFGELLAELKIPVYSLRSPGYAI 180
Db 120 DILKRCFCDIYVSNKKMLKQESRPDYVADAVFPFGELLAELKIPVYSLRSPGYAI 179
Qy 181 EKHSQGLLPPSPYVPMVSELSDOMTFIERVKNMIVLYFEFVFOIFDMKKMDQFYSEV 240
Db 180 EKHSQGLLPPSPYVPMVSELSDOMTFIERVKNMIVLYFEFVFOIFDMKKMDQFYSEV 239
Qy 241 GRPTTLSETMAKADIWLIRNYWDFPRLPLPNVEFVGLHCKPAKPLPKMEEFVQSS 300
Db 240 GRPTTLSETMAKADIWLIRNYWDFPRLPLPNVEFVGLHCKPAKPLPKMEEFVQSS 299
Qy 301 ENGIVVSLGSMVNTSEERANVIALAKIPQVLMRFDGNKPDGLNTRLYKMIPO 360
Db 300 ENGIVVSLGSMVNTSEERANVIALAKIPQVLMRFDGNKPDGLNTRLYKMIPO 359
Qy 361 DILGHPKTAFTTHGGMGIYEALYHGVPMVGPVIFGDDLDNIAMKAKAAVEINFTM 420
Db 360 DILGHPKTAFTTHGGMGIYEALYHGVPMVGPVIFGDDLDNIAMKAKAAVEINFTM 419
Qy 421 TSEDLRALRTVITDSSYKENAMRLSRHHDPVYKPLDRAVFWIEFVMBHKGAKLRSA 480
Db 420 TSEDLRALRTVITDSSYKENAMRLSRHHDPVYKPLDRAVFWIEFVMBHKGAKLRSA 479
Qy 481 HDLTFWQYHSLDVIGFLACVATVFIITKCELCFCFMKPAKGGKKRD 530
Db 480 HDLTFWQYHSLDVIGFLACVATVFIITKCELCFCFMKPAKGGKKRD 529

RESULT 6

S11309

glucuronosyltransferase (EC 2.4.1.17) - human
N/Alternate names: estradiol UDPglucuronosyltransferase
C/Species: Homo sapiens (man)
C/Date: 21-Nov-1993 #sequence_revision 10-Nov-1995 #text_change 29-Sep-1999

C/Accession: S11309
 R/Coffman, B.L.; Tephly, T.R.; Irshaid, Y.M.; Green, M.D.; Smith, C.; Jackson, M.R.; Moc Arch. Biochem. Biophys. 261, 170-175, 1990
 A/Title: Characterization and primary sequence of a human hepatic microsomal estriol UDP A/Reference number: S11309; MUID:90343358; PMID:2116769
 A/Accession: S11309
 A/Status: preliminary
 A/Molecule type: mRNA
 A/Residues: 1-523 <COP>
 A/Cross-references: GB:U06641; NID:g458398; PIDN:AAA83406.1; PID:g458399
 A/Note: the authors translated the codon AGG for residue 412 as Thr and ACC for residue C/Suprafamily: glucuronosyltransferase
 C/Keywords: glycosyltransferase; hexosyltransferase; transmembrane protein

Query Match 74.6%; Score 2086.5; DB 2; Length 523;
 Best Local Similarity 73.9%; Pred. No. 5.5e-153;
 Matches 385; Conservative 57; Mismatches 78; Indels 1; Gaps 1;

Qy 10 LLIQLSCYFSSGSCGKVLVWPTFEFHHNNIKITIDELVQGHVTVLASSASISFDPSP 69
 Db 3 LLIQLSCYFSSGSCGKVLVWPTFEFHHNNIKITIDELVQGHVTVLASSASISFDPSP 62
 Qy 70 STKEFVYVPSLTKEFEEDIKQLVKRVA-ELPKDTFWSYFSQVQEIWMTNDILRKECK 128
 Db 63 SAIKLEVPYPTSLTKNYLDBSLKIDRWIYGVSKTFWSYFSQVQELCMAYVDYENKLCX 122
 Qy 129 DIVSKTKMLKQLQESRPVVLADAFPPGELLAELKIPVYSLSRFSFGYAEKSGGL 188
 Db 123 DAVLNKTKMLKQLQESRPVVLADAFPPGELLAELKIPVYSLSRFSFGYAEKSGGL 182
 Qy 189 FPPSVYVPMSELSDQMTFIERVKMIVLYFEFQIYFDKMKMQDFSEVLGRFTTSE 248
 Db 183 FPPSVYVPMSELSDQMTFIERVKMIVLYFEFQIYFDKMKMQDFSEVLGRFTTSE 242
 Qy 249 TMAKADIWLIRNYWDFOPHPLLPVVEFVGGLHCKPAKPLPKMEEFVQSSGEGNVVFS 308
 Db 243 TMGKAEMLLIRNYWDFOPHPLLPVVEFVGGLHCKPAKPLPKMEEFVQSSGEGNVVFS 302
 Qy 309 LGSWVSNSEERANVIAALAKIPQKVLMPGKNDPLGLNTRLYKMTIPONDILGHKPT 368
 Db 303 LGSWVSNSEERANVIAALAKIPQKVLMPGKNDPLGLNTRLYKMTIPONDILGHKPT 362
 Qy 369 KAFITGGMNGIYEAIYHGVPMVGPVIFGDDLDNIJAHKAKAQAIVEINFTTSSEDLRA 428
 Db 363 KAFITGGMNGIYEAIYHGVPMVGPVIFGDDLDNIJAHKAKAQAIVEINFTTSSEDLRA 422
 Qy 429 LRTVITDSSYKENAMRLSRHHDDQVKKPLDRAVFWIEFVRRHKGAKHLRSAADLTWFOH 488
 Db 423 LKSVINDVYKENAMRLSRHHDDQVKKPLDRAVFWIEFVRRHKGAKHLRSAADLTWFOH 482
 Qy 489 YSIDVIGFLTCVATAIFLTGCFPLFSCQKPKTKIKRE 529
 Db 483 HSLDVIAFLAMVATVIFITKFCFLCFRKLAKKKKKRD 523

RESULT 7
 C/Accession: S11309
 R/Coffman, B.L.; Tephly, T.R.; Irshaid, Y.M.; Green, M.D.; Smith, C.; Jackson, M.R.; Moc Arch. Biochem. Biophys. 261, 170-175, 1990
 A/Title: Characterization and primary sequence of a human hepatic microsomal estriol UDP A/Reference number: S11309; MUID:90343358; PMID:2116769
 A/Accession: S11309
 A/Status: preliminary
 A/Molecule type: mRNA
 A/Residues: 1-530 <TVK>
 A/Cross-references: GB:L01082; NID:g165798; PIDN:AAA18021.1; PID:g165799
 C/Suprafamily: glucuronosyltransferase
 C/Keywords: glycosyltransferase; hexosyltransferase; transmembrane protein

Query Match 71.2%; Score 1991.5; DB 2; Length 530;
 Best Local Similarity 67.9%; Pred. No. 1.2e-145;
 Matches 360; Conservative 79; Mismatches 90; Indels 1; Gaps 1;

Qy 1 MSKMTSA-LIIQLSCYFSSGSCGKVLVWPTFEFHHNNIKITIDELVQGHVTVLASS 59
 Db 1 MSKMTSA-LIIQLSCYFSSGSCGKVLVWPTFEFHHNNIKITIDELVQGHVTVLASS 59
 Qy 60 ASISPDNSPSTLKEFVYVPSLTKEFEEDIKQLVKRVAELPKDTFWSYFSQVQEIWMTF 119
 Db 61 ASISPDNSPSTLKEFVYVPSLTKEFEEDIKQLVKRVAELPKDTFWSYFSQVQEIWMTF 120
 Qy 120 NDILRKECDIVSKTKMLKQLQESRPVVLADAFPPGELLAELKIPVYSLSRFSFGYAE 179
 Db 121 SDSCENACKEVYFMTKQLQESRPVVLADAFPPGELLAELKIPVYSLSRFSFGYAE 180
 Qy 180 IEKSGGLFPPSVYVPMSELSDQMTFIERVKMIVLYFEFQIYFDKMKMQDFSEV 239
 Db 181 MEKSGGLFPPSVYVPMSELSDQMTFIERVKMIVLYFEFQIYFDKMKMQDFSEV 240
 Qy 240 LGRFTTSETMAKADIWLIRNYWDFOPHPLLPVVEFVGGLHCKPAKPLPKMEEFVQSS 299
 Db 241 LGRFTTSETMAKADIWLIRNYWDFOPHPLLPVVEFVGGLHCKPAKPLPKMEEFVQSS 300
 Qy 300 GENGWVFSLSGWSNTEERANVIAALAKIPQKVLMPGKNDPLGLNTRLYKMTIPQ 359
 Db 301 GENGWVFSLSGWSNTEERANVIAALAKIPQKVLMPGKNDPLGLNTRLYKMTIPQ 360
 Qy 360 NDILGHKPTKAFVTHGANGIYEAIYHGVPMVGPVIFGDDLDNIJAHKAKAQAIVEINFT 419
 Db 361 NDILGHKPTKAFVTHGANGIYEAIYHGVPMVGPVIFGDDLDNIJAHKAKAQAIVEINFT 420
 Qy 420 MTSDDLALRPTVITDSSYKENAMRLSRHHDDQVKKPLDRAVFWIEFVRRHKGAKHLRSA 479
 Db 421 MTSDDLALRPTVITDSSYKENAMRLSRHHDDQVKKPLDRAVFWIEFVRRHKGAKHLRSA 480
 Qy 480 AHDLTWFOHYSIDVIGFLTCVATAIFLTGCFPLFSCQKPKTKIKRE 529
 Db 481 AHDLTWFOHYSIDVIGFLTCVATAIFLTGCFPLFSCQKPKTKIKRE 530

RESULT 8
 C/Accession: S11309
 R/Coffman, B.L.; Tephly, T.R.; Irshaid, Y.M.; Green, M.D.; Smith, C.; Jackson, M.R.; Moc Arch. Biochem. Biophys. 261, 170-175, 1990
 A/Title: Characterization and primary sequence of a human hepatic microsomal estriol UDP A/Reference number: S11309; MUID:90343358; PMID:2116769
 A/Accession: S11309
 A/Status: preliminary
 A/Molecule type: DNA
 A/Residues: 1-529 <MAC>
 A/Cross-references: GB:U05482
 A/Note: the authors translated the codon GTA for residue 57 as B, and GNC for residue 11 R/Mackenzie, P.I.
 J. Biol. Chem. 261, 6119-6125, 1986
 A/Title: Rat liver UDP-glucuronosyltransferase. Sequence and expression of a cDNA encodi A/Reference number: A24324; MUID:86196018; PMID:3084479
 A/Accession: A24324
 A/Status: preliminary
 A/Molecule type: mRNA
 A/Residues: 1-407, 'V', 409-529 <MA2>
 A/Cross-references: GB:M3506; NID:g207580; PIDN:AAA42313.1; PID:g207581
 A/Experimental source: liver
 C/Suprafamily: glucuronosyltransferase
 C/Keywords: glycosyltransferase; hexosyltransferase; transmembrane protein

Query Match 70.6%; Score 1975.5; DB 2; Length 529;
 Best Local Similarity 68.4%; Pred. No. 2e-144;
 Matches 361; Conservative 68; Mismatches 98; Indels 1; Gaps 1;

```

QY 1 MSKMTSALLLIQLSCYFSSGCGKVLWPTFESHMNIKTILDELVQRGHEVTVLASSA 60
DB 1 MSKMTSALLLIQLSCYFSSGCGKVLWPTFESHMNIKTILDELVQRGHEVTVLASSA 60
QY 61 SISFDPSPTLKEEYVAVSLTKEFEDIIKQVKNRA-ELPKOTFMSYSQVOEIMWTF 119
DB 61 SISFDPSPTLKEEYVAVSLTKEFEDIIKQVKNRA-ELPKOTFMSYSQVOEIMWTF 119
QY 120 NDILRKCKDVSXKMLMKKQLQESRPDVVLADAVFPFEGELIAELIKIPVYSLRFSFGYA 179
DB 120 NDILRKCKDVSXKMLMKKQLQESRPDVVLADAVFPFEGELIAELIKIPVYSLRFSFGYA 179
QY 121 SDVVENICKALIMKSLMKKQLQESRPDVVLADAVFPFEGELIAELIKIPVYSLRFSFGYA 180
DB 121 SDVVENICKALIMKSLMKKQLQESRPDVVLADAVFPFEGELIAELIKIPVYSLRFSFGYA 180
QY 180 IEKHSGLLPSPSYVAVVMSLSDQMTFIERVKMNIYVLYFEFWFOJFMKKMDQFYSV 239
DB 180 IEKHSGLLPSPSYVAVVMSLSDQMTFIERVKMNIYVLYFEFWFOJFMKKMDQFYSV 239
QY 181 CEKSGGLPSPSYVAVVMSLSDQMTFIERVKMNIYVLYFEFWFOJFMKKMDQFYSV 240
DB 181 CEKSGGLPSPSYVAVVMSLSDQMTFIERVKMNIYVLYFEFWFOJFMKKMDQFYSV 240
QY 240 LGRTTILSETMAKADIMLRNYWDFOPPHPLPNVEFVYGLHCKPAKPLPKMEEFVQSS 299
DB 240 LGRTTILSETMAKADIMLRNYWDFOPPHPLPNVEFVYGLHCKPAKPLPKMEEFVQSS 299
QY 241 LGRTTILSETMAKADIMLRNYWDFOPPHPLPNVEFVYGLHCKPAKPLPKMEEFVQSS 300
DB 241 LGRTTILSETMAKADIMLRNYWDFOPPHPLPNVEFVYGLHCKPAKPLPKMEEFVQSS 300
QY 300 GENGAVVFSLSGMSVNTSEERANVIASALAKIPQKVLMPFDGKPKDPTLGINTRLYKMI 359
DB 300 GENGAVVFSLSGMSVNTSEERANVIASALAKIPQKVLMPFDGKPKDPTLGINTRLYKMI 359
QY 301 GENGAVVFSLSGMSVNTSEERANVIASALAKIPQKVLMPFDGKPKDPTLGINTRLYKMI 360
DB 301 GENGAVVFSLSGMSVNTSEERANVIASALAKIPQKVLMPFDGKPKDPTLGINTRLYKMI 360
QY 360 NDILGHPKTKAFITGCGMNGIYEAIVGVPMVGPPIFGDQDNIAMKAKGAIVEINFKT 419
DB 360 NDILGHPKTKAFITGCGMNGIYEAIVGVPMVGPPIFGDQDNIAMKAKGAIVEINFKT 419
QY 361 NDILGHPKTKAFITGCGMNGIYEAIVGVPMVGPPIFGDQDNIAMKAKGAIVEINFKT 420
DB 361 NDILGHPKTKAFITGCGMNGIYEAIVGVPMVGPPIFGDQDNIAMKAKGAIVEINFKT 420
QY 420 MTSEDLRLPRTVITDSSYKENAMRLSRHHQPVKPLDRAVFIEMVRHKGAKHLRSA 479
DB 420 MTSEDLRLPRTVITDSSYKENAMRLSRHHQPVKPLDRAVFIEMVRHKGAKHLRSA 479
QY 421 LSTGLLTALAKIVNNDPSYKENAMRLSRHHQPVKPLDRAVFIEMVRHKGAKHLRSA 480
DB 421 LSTGLLTALAKIVNNDPSYKENAMRLSRHHQPVKPLDRAVFIEMVRHKGAKHLRSA 480
QY 480 AHDLTWFOHYSIDVIGFLTCVATAIFLTCKCFESQCKFNKTRKIEKE 527
DB 480 AHDLTWFOHYSIDVIGFLTCVATAIFLTCKCFESQCKFNKTRKIEKE 527
QY 481 LHDLSMFOHYSIDVIGFLTCVATAIFLTCKCFESQCKFNKTRKIEKE 528
DB 481 LHDLSMFOHYSIDVIGFLTCVATAIFLTCKCFESQCKFNKTRKIEKE 528

```

RESULT 9

```

B47113
glucuronosyltransferase (EC 2.4.1.17) UGT2B13 precursor - rabbit
C/Species: Oryctolagus cuniculus (domestic rabbit)
C/Date: 03-May-1994 #sequence_revision 03-May-1994 #text_change 29-Sep-1999
C/Accession: B47113
R/Title: Cloning and characterization of rabbit liver UDP-glucuronosyltransferase cDNAs
R/Key: R.H.; Pendurthi, U.R.; Nguyen, N.T.; Green, M.D.; Tephly, T.R.
J. Biol. Chem. 268, 15260-15266, 1993
A/Title: Cloning and characterization of rabbit liver UDP-glucuronosyltransferase cDNAs
A/Reference number: A47113; MUID:9315511; PMID:8325897
A/Accession: B47113
A/Status: preliminary
A/Molecule type: mRNA
A/Residues: 1-531 <TRK>
A/Cross-references: GB:L01061; NID:g165796; PIDN:AAA18020.1; PID:g165797
C/Superfamily: glucuronosyltransferase
C/Keywords: glucosyltransferase; hexosyltransferase; transmembrane protein

```

```

Query Match 69.3%; Score 1938; DB 2; Length 531;
Best Local Similarity 68.7%; Pred. No. 1.6e-141;
Matches 365; Conservative 62; Mismatches 102; Indels 2; Gaps 2;
QY 1 MSKMTSALLLIQLSCYFSSGCGKVLWPTFESHMNIKTILDELVQRGHEVTVLASSA 59
DB 1 MSKMTSALLLIQLSCYFSSGCGKVLWPTFESHMNIKTILDELVQRGHEVTVLASSA 60
QY 60 ASISFDPSPTLKEEYVAVSLTKEFEDIIKQVKNRA-ELPKOTFMSYSQVOEIMWTF 118
DB 60 ASISFDPSPTLKEEYVAVSLTKEFEDIIKQVKNRA-ELPKOTFMSYSQVOEIMWTF 118
QY 61 ASISFDPSPTLKEEYVAVSLTKEFEDIIKQVKNRA-ELPKOTFMSYSQVOEIMWTF 120
DB 61 ASISFDPSPTLKEEYVAVSLTKEFEDIIKQVKNRA-ELPKOTFMSYSQVOEIMWTF 120
QY 119 FNDILRKCKDVSXKMLMKKQLQESRPDVVLADAVFPFEGELIAELIKIPVYSLRFSFGYA 178
DB 119 FNDILRKCKDVSXKMLMKKQLQESRPDVVLADAVFPFEGELIAELIKIPVYSLRFSFGYA 180
QY 121 YSDICBICKEVILMKKQLQESRPDVVLADAVFPFEGELIAELIKIPVYSLRFSFGYA 180
DB 121 YSDICBICKEVILMKKQLQESRPDVVLADAVFPFEGELIAELIKIPVYSLRFSFGYA 180
QY 179 AIEKHSGLLPSPSYVAVVMSLSDQMTFIERVKMNIYVLYFEFWFOJFMKKMDQFYSV 238
DB 179 AIEKHSGLLPSPSYVAVVMSLSDQMTFIERVKMNIYVLYFEFWFOJFMKKMDQFYSV 238

```

```

DB 181 MLOKGGGLLPSPSYVAVVMSLSDQMTFIERVKMNIYVLYFEFWFOJFMKKMDQFYSV 240
QY 239 VLRPPTLSETMAKADIMLRNYWDFOPPHPLPNVEFVYGLHCKPAKPLPKMEEFVQSS 298
DB 239 VLRPPTLSETMAKADIMLRNYWDFOPPHPLPNVEFVYGLHCKPAKPLPKMEEFVQSS 298
QY 241 VLRPPTLSETMAKADIMLRNYWDFOPPHPLPNVEFVYGLHCKPAKPLPKMEEFVQSS 300
DB 241 VLRPPTLSETMAKADIMLRNYWDFOPPHPLPNVEFVYGLHCKPAKPLPKMEEFVQSS 300
QY 299 GENGAVVFSLSGMSVNTSEERANVIASALAKIPQKVLMPFDGKPKDPTLGINTRLYKMI 358
DB 299 GENGAVVFSLSGMSVNTSEERANVIASALAKIPQKVLMPFDGKPKDPTLGINTRLYKMI 358
QY 301 GENGAVVFSLSGMSVNTSEERANVIASALAKIPQKVLMPFDGKPKDPTLGINTRLYKMI 360
DB 301 GENGAVVFSLSGMSVNTSEERANVIASALAKIPQKVLMPFDGKPKDPTLGINTRLYKMI 360
QY 359 QNDILGHPKTKAFITGCGMNGIYEAIVGVPMVGPPIFGDQDNIAMKAKGAIVEINFKT 418
DB 359 QNDILGHPKTKAFITGCGMNGIYEAIVGVPMVGPPIFGDQDNIAMKAKGAIVEINFKT 418
QY 361 QNDILGHPKTKAFITGCGMNGIYEAIVGVPMVGPPIFGDQDNIAMKAKGAIVEINFKT 420
DB 361 QNDILGHPKTKAFITGCGMNGIYEAIVGVPMVGPPIFGDQDNIAMKAKGAIVEINFKT 420
QY 419 TMSADLRLPRTVITDSSYKENAMRLSRHHQPVKPLDRAVFIEMVRHKGAKHLRSA 478
DB 419 TMSADLRLPRTVITDSSYKENAMRLSRHHQPVKPLDRAVFIEMVRHKGAKHLRSA 478
QY 479 AHDLTWFOHYSIDVIGFLTCVATAIFLTCKCFESQCKFNKTRKIEKE 529
DB 479 AHDLTWFOHYSIDVIGFLTCVATAIFLTCKCFESQCKFNKTRKIEKE 529
QY 481 AHDLTWFOHYSIDVIGFLTCVATAIFLTCKCFESQCKFNKTRKIEKE 531
DB 481 AHDLTWFOHYSIDVIGFLTCVATAIFLTCKCFESQCKFNKTRKIEKE 531

```

RESULT 10

```

S68200
glucuronosyltransferase (EC 2.4.1.17) precursor - rat
N/Alternate names: UDP-glucuronosyltransferase
C/Species: Rattus norvegicus (Norway rat)
C/Date: 15-Feb-1997 #sequence_revision 13-Mar-1997 #text_change 17-Nov-2000
C/Accession: S68200; A61266
R/Title: Cloning and expression of a rat liver phenobarbital-inducible UDP-glucuronosyltransferase
R/Key: R.H.; Pendurthi, U.R.; Nguyen, N.T.; Green, M.D.; Tephly, T.R.
J. Biol. Chem. 272, 460-468, 1997
A/Title: Cloning and expression of a rat liver phenobarbital-inducible UDP-glucuronosyltransferase
A/Reference number: S68200; MUID:96032669; PMID:7514722
A/Accession: S68200
A/Status: preliminary
A/Molecule type: mRNA
A/Residues: 1-530 <GRE>
A/Cross-references: EMBL:U06273; NID:g458394; PIDN:AAA83404.1; PID:g458395
R/Title: Cloning and expression of a rat liver phenobarbital-inducible UDP-glucuronosyltransferase
R/Key: R.H.; Pendurthi, U.R.; Nguyen, N.T.; Green, M.D.; Tephly, T.R.
J. Biol. Chem. 272, 460-468, 1997
A/Title: Cloning and expression of a rat liver phenobarbital-inducible UDP-glucuronosyltransferase
A/Reference number: A61266; MUID:9312327; PMID:1906977
A/Accession: A61266
A/Status: preliminary
A/Molecule type: protein
A/Residues: 24-38 <STR>
C/Superfamily: glucuronosyltransferase
C/Keywords: glycoprotein; glycosyltransferase; hexosyltransferase
F:1-23/Domain: signal sequence #stratus predicted <SIG>
F:24-530/Product: glucuronosyltransferase #stratus predicted <MAT>

```

```

Query Match 69.0%; Score 1930.5; DB 2; Length 530;
Best Local Similarity 66.8%; Pred. No. 6e-141;
Matches 354; Conservative 73; Mismatches 102; Indels 1; Gaps 1;
QY 1 MSKMTSALLLIQLSCYFSSGCGKVLWPTFESHMNIKTILDELVQRGHEVTVLASSA 60
DB 1 MSKMTSALLLIQLSCYFSSGCGKVLWPTFESHMNIKTILDELVQRGHEVTVLASSA 60
QY 61 SISFDPSPTLKEEYVAVSLTKEFEDIIKQVKNRA-ELPKOTFMSYSQVOEIMWTF 119
DB 61 SISFDPSPTLKEEYVAVSLTKEFEDIIKQVKNRA-ELPKOTFMSYSQVOEIMWTF 119
QY 61 FVLDPKETSDDLKFTVFPFISFSHDLNFTFRVNVNTYELPDTCLSTYLYLQDTIDEX 120
DB 61 FVLDPKETSDDLKFTVFPFISFSHDLNFTFRVNVNTYELPDTCLSTYLYLQDTIDEX 120
QY 120 NDILRKCKDVSXKMLMKKQLQESRPDVVLADAVFPFEGELIAELIKIPVYSLRFSFGYA 179
DB 120 NDILRKCKDVSXKMLMKKQLQESRPDVVLADAVFPFEGELIAELIKIPVYSLRFSFGYA 179
QY 121 SDYCLAVCKEAVSNQKMTLQESKFDVPSDAIGCGELIAELIKIPVYSLRFSFGYT 180
DB 121 SDYCLAVCKEAVSNQKMTLQESKFDVPSDAIGCGELIAELIKIPVYSLRFSFGYT 180
QY 180 IEKHSGLLPSPSYVAVVMSLSDQMTFIERVKMNIYVLYFEFWFOJFMKKMDQFYSV 239
DB 180 IEKHSGLLPSPSYVAVVMSLSDQMTFIERVKMNIYVLYFEFWFOJFMKKMDQFYSV 239
QY 181 IEKHSGLLPSPSYVAVVMSLSDQMTFIERVKMNIYVLYFEFWFOJFMKKMDQFYSV 240
DB 181 IEKHSGLLPSPSYVAVVMSLSDQMTFIERVKMNIYVLYFEFWFOJFMKKMDQFYSV 240
QY 240 LGRTTILSETMAKADIMLRNYWDFOPPHPLPNVEFVYGLHCKPAKPLPKMEEFVQSS 299
DB 240 LGRTTILSETMAKADIMLRNYWDFOPPHPLPNVEFVYGLHCKPAKPLPKMEEFVQSS 299

```

```

Db 241 LGRPTTLAEIMGKEMWILRSWDELFPHISPNDVYIGLHCKRAKPLPDIEDFVQSS 300
Qy 300 GENGWVVSLSGSMVSNTESEERANVYASALAKIPOKVLNFRPGNKEDTIGLTRLYKMIPO 359
Db 301 GEHGVVVSLSGSMVSNTESEERANVYASALAKIPOKVLNFRPGNKEDTIGLTRLYKMIPO 360
Qy 360 NDLLGHPRTKAFITGKNGNGIYEALYHGVPMVGVIFPGDQDNINAHMKAKGAAYEINFKT 419
Db 361 NDLLGHPRTKAFITGKNGNGIYEALYHGVPMVGVIFPGDQDNINAHMKAKGAAYEINFKT 420
Qy 420 MTSBDLRLALRTVITDSSYKENAMRLSRIHHDQVKKPLDRAVFWIEFWRHKGAGHLSRA 479
Db 421 MTSBDLRLALRTVITDSSYKENAMRLSRIHHDQVKKPLDRAVFWIEFWRHKGAGHLSRA 480
Qy 480 AHDLTWFOHYSIDVIGFLITCVATAIFLFTKCFELFSCQKFNKTRKIEKRE 529
Db 481 GHNLPWQYHSLDVIGFLILSCVATVVALKCFELFVFRFVKKEKKTINE 530

```

RESULT 11

```

500163
glucuronosyltransferase (EC 2.4.1.17) precursor - mouse
N:Alternate names: UDP-glucuronosyltransferase
C:Species: Mus musculus (house mouse)
C:Date: 28-Aug-1989 #sequence_revision 28-Aug-1989 #text_change 29-Sep-1999
C:Accession: S00163
R:Kimura, T.; Owens, I.S.
Eur. J. Biochem. 168, 515-521, 1987
A:Title: Mouse UDP-glucuronosyltransferase. cDNA and complete amino acid sequence and re
A:Reference number: S00163; MUID:88029469; PMID:3117546
A:Accession: S00163
A:Molecule type: mRNA
A:Residues: 1-530 <KIM>
A:Cross-references: EMBL:X06358; NID:955119; PID:CAA29657.1; PID:955120
C:Comment: This enzyme catalyzes the conjugation of lipophilic compounds with glucuronic
C:Superfamily: glucuronosyltransferase
C:Keywords: glycoprotein; glycosyltransferase; hexosyltransferase; transmembrane protein
F:1-16/Domain: signal sequence #status predicted <SIG>
F:17-530/Product: glucuronosyltransferase #status predicted <MAT>
F:1494-510/Domain: transmembrane #status predicted <TM>
F:316/483/Binding site: carbohydrate (Asn) (covalent) #status predicted

```

```

Query Match 68.1%; Score 1904.5; DB 2; Length 530;
Best Local Similarity 66.8%; Pred. No. 6.1e-139;
Matches 354; Conservative 67; Mismatches 108; Indels 1; Gaps 1;

Qy 1 MSMKWTSAALLLIQLSCYSSSGCGVLMPTFESHMNIKITLDELVORGHEVTVLASSA 60
Db 1 MPMKWTSAALLLIQLSCYSSSGCGVLMPTFESHMNIKITLDELVORGHEVTVLASSA 60
Qy 61 SISFPNPSSTLKFEVYVSLTKTEFEDIKQVLRMA-ELPKDTFMSYFSOVQOIMWTF 119
Db 61 YVLDLPKKSPPGLKFTFPTSVSKDNLBNFIFKVDVMTYEMBRDCLSYPLQMMIDBF 120
Qy 120 NDILRPFCKDIYSNKKMKLQESRFDVVLADAVFPFGELLAEILKIPVYSLRPSGYA 179
Db 120 NDILRPFCKDIYSNKKMKLQESRFDVVLADAVFPFGELLAEILKIPVYSLRPSGYA 179
Qy 180 IEKHSGLLPPSPYVPMVSELSDQMTFIERVKNMIYVLYFEFPOIDPMKKMDQFYSYV 239
Db 180 IEKHSGLLPPSPYVPMVSELSDQMTFIERVKNMIYVLYFEFPOIDPMKKMDQFYSYV 239
Qy 241 LGRPTTLAEIMGKEMWILRSWDELFPHISPNDVYIGLHCKRAKPLPKMEEFVQSS 300
Db 241 LGRPTTLAEIMGKEMWILRSWDELFPHISPNDVYIGLHCKRAKPLPKMEEFVQSS 300
Qy 300 GENGWVVSLSGSMVSNTESEERANVYASALAKIPOKVLNFRPGNKEDTIGLTRLYKMIPO 359
Db 300 GENGWVVSLSGSMVSNTESEERANVYASALAKIPOKVLNFRPGNKEDTIGLTRLYKMIPO 359
Qy 360 NDLLGHPRTKAFITGKNGNGIYEALYHGVPMVGVIFPGDQDNINAHMKAKGAAYEINFKT 419
Db 360 NDLLGHPRTKAFITGKNGNGIYEALYHGVPMVGVIFPGDQDNINAHMKAKGAAYEINFKT 419

```

```

Db 361 NDLLGHPRTKAFITGKNGNGIYEALYHGVPMVGVIFPGDQDNINAHMKAKGAAYEINFKT 420
Qy 420 MTSBDLRLALRTVITDSSYKENAMRLSRIHHDQVKKPLDRAVFWIEFWRHKGAGHLSRA 479
Db 421 MTSBDLRLALRTVITDSSYKENAMRLSRIHHDQVKKPLDRAVFWIEFWRHKGAGHLSRA 480
Qy 480 AHDLTWFOHYSIDVIGFLITCVATAIFLFTKCFELFSCQKFNKTRKIEKRE 529
Db 481 GHNLPWQYHSLDVIGFLILSCVATVVALKCFELFVFRFVKKEKKTINE 530

```

RESULT 12

```

507390
glucuronosyltransferase (EC 2.4.1.17) 3 precursor - rat
N:Alternate names: 17beta-hydroxysteroid UDP-glucuronosyltransferase; UDP-glucuronosyltr
C:Species: Rattus norvegicus (Norway rat)
C:Date: 30-Sep-1991 #sequence_revision 30-Sep-1991 #text_change 07-May-1999
C:Accession: S07390; A33236; A28460; S59626
R:Harding, D.; Wilson, S.M.; Jackson, M.R.; Burchell, B.; Green, M.D.; Tephly, T.R.
Nucleic Acids Res. 15, 3936, 1987
A:Title: Nucleotide and deduced amino acid sequence of rat liver 17beta-hydroxysteroid l
A:Reference number: S07390; MUID:87231096; PMID:3108864
A:Accession: S07390
A:Molecule type: mRNA
A:Residues: 1-530 <HAR>
A:Cross-references: EMBL:Y00156
A:Experimental source: liver
A:Accession: A33236
A:Molecule type: protein
A:Residues: 24-61 <HAR2>
A:Experimental source: liver
R:Mackenzie, P.I.
J. Biol. Chem. 262, 9744-9749, 1987
A:Title: Rat liver UDP-glucuronosyltransferase. Identification of cDNAs encoding two enz
A:Reference number: A28460; MUID:87250645; PMID:3110162
A:Accession: A28460
A:Molecule type: mRNA
A:Residues: 1-118, 'G', 120-240, 'L', 242-423, 'S', 425-499, 'T', 501-530 <MAC>
A:Experimental source: liver
R:Yamashta, A.; Watanabe, M.; Tonegawa, T.; Sugitara, T.; Waku, K.
Biochem. J. 312, 301-308, 1995
A:Title: Acyl-CoA binding and acylation of UDP-glucuronosyltransferase isoforms of rat l
A:Reference number: S59626; MUID:96077159; PMID:7492328
A:Accession: S59626
A:Molecule type: protein
A:Residues: 24-44 <YAM>
A:Experimental source: liver
C:Superfamily: glucuronosyltransferase
C:Keywords: glycosyltransferase; hexosyltransferase; transmembrane protein
F:1-23/Domain: signal sequence #status predicted <SIG>
F:24-530/Product: glucuronosyltransferase #status experimental <MAT>
F:1494-510/Domain: transmembrane #status predicted <TM>

```

```

Query Match 67.6%; Score 1892.5; DB 2; Length 530;
Best Local Similarity 65.8%; Pred. No. 5.2e-138;
Matches 349; Conservative 70; Mismatches 110; Indels 1; Gaps 1;

Qy 1 MSMKWTSAALLLIQLSCYSSSGCGVLMPTFESHMNIKITLDELVORGHEVTVLASSA 60
Db 1 MPMKWTSAALLLIQLSCYSSSGCGVLMPTFESHMNIKITLDELVORGHEVTVLASSA 60
Qy 61 SISFPNPSSTLKFEVYVSLTKTEFEDIKQVLRMA-ELPKDTFMSYFSOVQOIMWTF 119
Db 61 YVLDLPKKSPPGLKFTFPTSVSKDNLBNFIFKVDVMTYEMBRDCLSYPLQMMIDBF 120
Qy 120 NDILRPFCKDIYSNKKMKLQESRFDVVLADAVFPFGELLAEILKIPVYSLRPSGYA 179
Db 120 NDILRPFCKDIYSNKKMKLQESRFDVVLADAVFPFGELLAEILKIPVYSLRPSGYA 179
Qy 241 LGRPTTLAEIMGKEMWILRSWDELFPHISPNDVYIGLHCKRAKPLPKMEEFVQSS 300
Db 241 LGRPTTLAEIMGKEMWILRSWDELFPHISPNDVYIGLHCKRAKPLPKMEEFVQSS 300
Qy 300 GENGWVVSLSGSMVSNTESEERANVYASALAKIPOKVLNFRPGNKEDTIGLTRLYKMIPO 359
Db 300 GENGWVVSLSGSMVSNTESEERANVYASALAKIPOKVLNFRPGNKEDTIGLTRLYKMIPO 359
Qy 360 NDLLGHPRTKAFITGKNGNGIYEALYHGVPMVGVIFPGDQDNINAHMKAKGAAYEINFKT 419
Db 360 NDLLGHPRTKAFITGKNGNGIYEALYHGVPMVGVIFPGDQDNINAHMKAKGAAYEINFKT 419

```


[illegible]

Qy 240 LGRPTLSETMAKADIWILIRNVWDFQPHPLLPVNEFVGLHCKPAKPLPKMESEFVQSS 299
Db 241 LGRPTTLAETNGKEMWILIRSYWDFEFHPLPNDVYIGGLQCKPAKPLPDIDIDFVQSS 300
Qy 300 GENGWVFSLSGMSVNTSEEPANVASALAKIPOKVLWRFQGNKPDITGLNTRLYKMIPO 359
Db 301 GEHGVVFSLSGMSVNTSEEPANVASALAKIPOKVLWRFQGNKPDITGLNTRLYKMIPO 360
Qy 360 NDLLGHPKTKAFITGNGNGIYEALYHGVPMVGVPIFGDQDNLNIAHMKAKGAAVEINFKT 419
Db 361 NDLLGHPKTKAFITGNGNGIYEALYHGVPMVGVPIFGDQDNLNIAHMKAKGAAVEINFKT 420
Qy 420 MTSDDLALRTRVITDSSYKENAMLSRIHHDQPVKPLDRAVFWIEFVRRHKGAKHLRSA 479
Db 421 MSKSDLFALKEVINNPYKKNAMLSRIHHDQPVKPLDRAVFWIEFVRRHKGAKHLRSA 480
Qy 480 AHDLTWFOHYSIDVIGFLITCVATAIFLTKCFPLFCOKFNKTRKIEKRE 529
Db 481 GHNLPWYQYHSLDVIGFLITCVATAIFLTKCFPLFCOKFNKTRKIEKRE 530

RESULT 15

S15089
glucuronosyltransferase (EC 2.4.1.17) - rat
C:Species: Rattus norvegicus (Norway rat)
C:Date: 21-Nov-1993 #sequence_revision 10-Nov-1995 #text_change 20-Jun-2000
C:Accession: S15089
R:Laizard, D.; Zupko, K.; Portia, Y.; Neff, P.; Lazarovite, J.; Horn, S.; Khen, M.; Lancet, Nature 349, 790-793, 1991
A:Title: Odorant signal termination by olfactory UDP glucuronosyl transferase.
A:Reference number: S15089; MUID:91156050; PMID:1900353
A:Accession: S15089
A:Molecule type: mRNA
A:Residues: 1-527 <LAZ>
A:Cross-reference: GB:X57565; NID:957762; PIDN:CAA40797.1; PID:93980217
C:Superfamily: glucuronosyltransferase
C:Keywords: glycosyltransferase, hexosyltransferase

Query Match 62.9%; Score 1760.5; DB 2; Length 527;

Best Local Similarity 62.0%; Pred. No. 7.8e-128; Indels 9; Gaps 3;

Matches 327; Conservative 77; Mismatches 114; Indels 9; Gaps 3;

Qy 5 WTSALLLIQLSCYFSSGCGKVLVWPTFSHMMNIXITLDELVQRGHEVTVLASSASISF 64
Db 8 WSLQSLSLGMSL-----GNNVLWPMGSHWLVNKKIIDELRKHNVTVLVASGALFI 61
Qy 65 DPNSPSTLKFEVYPSVLTKEFEDIIKQIVKMAE--LPKDTFWSYFSQVOEIMTFENDI 122
Db 62 TPSVSPSLTFEIIYPPVPGKEKIESVIKDFVLTWLENRPSPTIMTFYKEMAKVIEFHLV 121
Qy 123 LRKFKDIVSNKKMLKLOESRFDVVLDAVFPGBELLAEILKIPVVSILRSPGATK 182
Db 122 SRGIDGVALKNKMTKLRGKFEVLLSDPVPCCDIVALKIGIPITSLRSPASTVEK 181
Qy 183 HSGGLLPSPSVVPMSEISDQMTFIERVKNMIYVLYFEFQIFDMKKMDQFSEVLGR 242
Db 182 HCGKVPFSPSVVPMSEISDQMTFIERVKNMIYVLYFEFQIFDMKKMDQFSEVLGR 240
Qy 243 PTTLSETMAKADIWILIRNVWDFQPHPLLPVNEFVGLHCKPAKPLPKMESEFVQSSGN 302
Db 241 PTTLSETMAKADIWILIRNVWDFQPHPLLPVNEFVGLHCKPAKPLPKMESEFVQSSGN 300
Qy 303 GVVVPSLSGMSVNTSEEPANVASALAKIPOKVLWRFQGNKPDITGLNTRLYKMIPO 362
Db 301 GVVVPSLSGMSVNTSEEPANVASALAKIPOKVLWRFQGNKPDITGLNTRLYKMIPO 360
Qy 363 LGHPKTKAFITGNGNGIYEALYHGVPMVGVPIFGDQDNLNIAHMKAKGAAVEINFKT 422
Db 361 LGHPKTKAFITGNGNGIYEALYHGVPMVGVPIFGDQDNLNIAHMKAKGAAVEINFKT 420
Qy 423 EDLLRALRTRVITDSSYKENAMLSRIHHDQPVKPLDRAVFWIEFVRRHKGAKHLRSA 482
Db 421 ADLLSAVRAVINPEPKENAMLSRIHHDQPVKPLDRAVFWIEFVRRHKGAKHLRSA 480

Qy 483 LTFWFOHYSIDVIGFLITCVATAIFLTKCFPLFCOKFNKTRKIEKRE 529
Db 481 LSWFQYHSLDVIGFLITCVATAIFLTKCFPLFCOKFNKTRKIEKRE 527

Search completed: December 5, 2003, 09:50:16
Job time: 22 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 5, 2003, 09:46:05 ; Search time 39 seconds
(without alignments)
3500.252 Million cell updates/sec

Title: US-09-980-729B-5
Perfect score: 2798
Sequence: 1 MSKMWSALLILQSCYFSS.....KCFLFSCQKFKTKIEKRE 529

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues
Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

SPTREMBL_23:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_virus:*
16: sp_bacterioph:*
17: sp_archaeoph:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|--------|-------------|--------|-------|-------------|
| 1 | 2368.5 | 84.6 | 528 | 6 | Q8WN97 |
| 2 | 2311 | 82.6 | 529 | 6 | Q97951 |
| 3 | 2291 | 81.9 | 529 | 6 | Q9GLD9 |
| 4 | 2289 | 81.8 | 529 | 6 | Q9GLD9 |
| 5 | 2261 | 80.8 | 529 | 6 | Q9TSL6 |
| 6 | 1999.5 | 71.5 | 529 | 11 | Q8R084 |
| 7 | 1993.5 | 71.2 | 532 | 11 | Q8K154 |
| 8 | 1970.5 | 70.4 | 528 | 11 | Q8VIF9 |
| 9 | 1908.5 | 68.2 | 530 | 11 | Q8K169 |
| 10 | 1895.5 | 67.7 | 530 | 11 | Q91WH2 |
| 11 | 1895 | 67.7 | 529 | 11 | Q8BJL9 |
| 12 | 1878 | 67.1 | 529 | 11 | Q8VIF8 |
| 13 | 1854.5 | 66.3 | 530 | 11 | Q8VCN3 |
| 14 | 1802 | 64.4 | 530 | 11 | Q9R110 |
| 15 | 1795.5 | 64.2 | 527 | 4 | Q9Y4X1 |
| 16 | 1780.5 | 63.6 | 527 | 11 | Q9ESB4 |

| | | | | | | |
|----|--------|------|-----|----|--------|--------------------|
| 17 | 1747.5 | 62.5 | 534 | 11 | Q8BWQ1 | Q8bwq1 mus musculu |
| 18 | 1739.5 | 62.2 | 534 | 11 | Q8RI29 | Q8ri29 mus musculu |
| 19 | 1737.5 | 62.1 | 534 | 11 | Q9DB11 | Q9db11 mus musculu |
| 20 | 1721 | 61.5 | 449 | 4 | Q9H6S4 | Q9h6s4 homo sapien |
| 21 | 1204.5 | 43.0 | 531 | 11 | Q8VD45 | Q8vd45 homo sapien |
| 22 | 1192 | 42.6 | 533 | 6 | Q46549 | Q46549 felis silve |
| 23 | 1144.5 | 40.9 | 533 | 6 | Q9TSL7 | Q9tsl7 macaca fasc |
| 24 | 1144.5 | 40.9 | 533 | 6 | Q9SKM4 | Q9skm4 macaca mula |
| 25 | 1138.5 | 40.7 | 533 | 6 | Q46423 | Q46423 felis silve |
| 26 | 1135.5 | 40.6 | 533 | 6 | Q9BDG7 | Q9bdg7 macaca mula |
| 27 | 1132.5 | 40.5 | 533 | 6 | Q46548 | Q46548 felis silve |
| 28 | 1121.5 | 40.1 | 530 | 6 | Q9TSM0 | Q9tsm0 macaca fasc |
| 29 | 1120.5 | 40.0 | 530 | 4 | Q9HAW9 | Q9haw9 homo sapien |
| 30 | 1120.5 | 40.0 | 530 | 4 | Q60656 | Q60656 homo sapien |
| 31 | 1116.5 | 39.9 | 530 | 4 | Q8TEX4 | Q8tex4 homo sapien |
| 32 | 1114.5 | 39.8 | 530 | 6 | Q9TSL8 | Q9tsl8 macaca fasc |
| 33 | 1113 | 39.8 | 529 | 6 | Q18736 | Q18736 bos taurus |
| 34 | 1102.5 | 39.4 | 530 | 4 | Q14928 | Q14928 homo sapien |
| 35 | 1101.5 | 39.4 | 530 | 4 | Q9HAW7 | Q9haw7 homo sapien |
| 36 | 1099.5 | 39.3 | 530 | 4 | Q9HAW8 | Q9haw8 homo sapien |
| 37 | 1092.5 | 39.0 | 531 | 6 | Q9XS55 | Q9xs55 ovis aries |
| 38 | 1089.5 | 38.9 | 530 | 4 | Q00473 | Q00473 homo sapien |
| 39 | 1088.5 | 38.9 | 530 | 11 | Q8VD44 | Q8vd44 ratcus norv |
| 40 | 1083 | 38.7 | 530 | 4 | Q00474 | Q00474 homo sapien |
| 41 | 1081.5 | 38.7 | 531 | 11 | Q8VD43 | Q8vd43 ratcus norv |
| 42 | 1081.5 | 38.7 | 532 | 4 | Q96TE7 | Q96te7 homo sapien |
| 43 | 1077 | 38.5 | 532 | 6 | Q9XS56 | Q9xs56 ovis aries |
| 44 | 1076.5 | 38.5 | 532 | 4 | Q8WUD4 | Q8wud4 homo sapien |
| 45 | 1075 | 38.4 | 498 | 11 | Q8VCQ9 | Q8vcq9 mus musculu |

ALIGNMENTS

| RESULT 1 | ID | Q8WN97 | PRELIMINARY; | PRT; | 528 AA. |
|-----------------------|---|------------------|---------------------|-----------|-------------|
| AC | Q8WN97/ | | | | |
| DT | 01-MAR-2002 (TReMBLrel. 20, Created) | | | | |
| DT | 01-MAR-2002 (TReMBLrel. 20, Last sequence update) | | | | |
| DT | 01-MAR-2003 (TReMBLrel. 23, Last annotation update) | | | | |
| DE | UDP-glucuronosyltransferase 2B30. | | | | |
| GN | UGT2B30. | | | | |
| OS | Macaca fascicularis (Cray eating macaque) (Cynomolpus monkey). | | | | |
| OC | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; | | | | |
| OC | Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea; | | | | |
| OC | Cercopithecoidea; Macaca. | | | | |
| OX | NCBI_TaxID=9541; | | | | |
| RN | [1] | | | | |
| RP | SEQUENCE FROM N.A. | | | | |
| RA | Girard C., Barbier O., Belanger A.; | | | | |
| RT | "Structure of the monkey UDP-glucuronosyltransferase UGT2B30 gene and | | | | |
| RT | characterization of the expressed enzyme which conjugates androgens, | | | | |
| RT | estrogens and progestins, in addition to cortisol and aldosterone | | | | |
| RT | metabolites." | | | | |
| RL | Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases. | | | | |
| DR | EMBL; AF401657; A160145.1; - | | | | |
| DR | InterPro; IPR002213; UDP_gluco_trans. | | | | |
| DR | Pfam; PF00201; UDPGT; 1. | | | | |
| DR | PROSITE; PS00375; UDPGT; 1. | | | | |
| KW | Transferase. | | | | |
| SO | SEQUENCE 528 AA; 60366 MW; 361349D4717C68D5 CRC64; | | | | |
| Query Match | | 84.6%; | Score 2368.5; | DB 6; | Length 528; |
| Best Local Similarity | | 83.2%; | Pred. No. 3.1e-180; | | |
| Matches 440; | | Conservative 40; | Mismatches 48; | Indels 1; | Gaps 1; |
| Qy | 1 MSKMWSALLILQSCYFSSGCGKVLWPTFSHMMNKTITIDELVORGHETVLASSA 60 | | | | |
| Db | 1 MSKMWSALLILQSCYFSSGCGKVLWPTFSHMMNKTITIDELVORGHETVLASSA 60 | | | | |
| Qy | 61 SIFDPNPSSTLKEFYVPSLTKTEFEDIKQLVKEVAELPKDTFWSYFQVOEIMWTFN 120 | | | | |

```
Db 61 SILRPNNSALKFEICPTSLTEFEDSVTLQVKNMSDIPKDETFWPHFLQVQEMMTYTG 120
Qy 121 DILKFCDDIVSNKKLMKKLQESRFDVVLADAVPFGGLLAELKIPVYSLRSPGYAI 180
Db 121 DMIRFCDDIVSNKKLMKKLQESRFDVVLADAIISCGELLAELKIPVYSLRSPGYAI 180
Qy 181 EKHSGLLFPSPYVPMVMSLSDDQMTFIERVKMNIYVLYFEFWQIPMKMKDQYSEVL 240
Db 181 EKHSGLLFPSPYVPMVMSLSDDQMTFIERVKMNIYVLYFEFWQIPMKMKDQYSEVL 240
Qy 241 GRPPTLSETMAKADIWILIRNWDFOFPHPLPNVEFVGLHCKPAKLPKMEEFVQSSG 300
Db 241 GRPPTLSETMAKADIWILIRNWDFOFPHPLPNVEFVGLHCKPAKLPKMEEFVQSSG 300
Qy 301 ENGVAVPSLGSVMSTSEERANVITASALAKIPQKVLMPDQKPTGLNTRLKYMPION 360
Db 301 ENGVAVPSLGSVMSTSEERANVITASALAKIPQKVLMPDQKPTGLNTRLKYMPION 360
Qy 361 DILGHPKTRAFITTHGNGNGIYEAIYHGVPMVGPPIFGDQDNIAHMKAKGAIVEINFKT 420
Db 361 DILGHPKTRAFITTHGNGNGIYEAIYHGVPMVGPPIFGDQDNIAHMKAKGAIVEINFKT 420
Qy 421 TSEDLRALRTVITDSSYKENAMRLSRIHDDQPKPLDRAVFWIEFWRHKGAGHLSAA 480
Db 421 TSEDLRALRTVITDSSYKENAMRLSRIHDDQPKPLDRAVFWIEFWRHKGAGHLSAA 480
Qy 481 HDLTFQYHSIDVIGFLTCVATAIFLTCKFLFSCCKFNKTRKIEKRE 529
Db 481 HDLTFQYHSIDVIGFLTCVATAIFLTCKFLFSCCKFNKTRKIEKRE 529
```

RESULT 2

```
097951 ID 097951 PRELIMINARY; PRT; 529 AA.
AC 097951;
DT 01-MAY-1999 (TREMBlrel. 10, Created)
DT 01-MAY-1999 (TREMBlrel. 10, Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE UDP-glucuronosyltransferase.
GN UGT2B18.
OS Macaca fascicularis (Crab eating macaque) (Synonym: J. monkey).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
OC Cercopithecinae; Macaca.
CX NCBI_TaxID=9541;
RN [1]
RP SEQUENCE FROM N.A.
RA Beaulieu M., Levesque E., Barbier O., Turgeon D., Belanger G.,
RA Hum D.W., Belanger A.;
RL Submitted (JUL-1997) to the EMBL/Genbank/DBJ databases.
DR EMBL; AF016310; AAC98726.1; -.
DR InterPro; IPR002213; UDP_gluco_trans.
DR Pfam; PF00201; UDPGT; 1.
DR PROSITE; PS00375; UDPGT; 1.
KW Transferase.
SQ SEQUENCE 529 AA; 60801 MW; 3CECB497B8C3601F CRC64;
```

Query Match 82.6%; Score 2311; DB 6; Length 529;
Best Local Similarity 81.3%; Pred. No. 1,2e-175;
Matches 430; Conservative 39; Mismatches 60; Indels 0; Gaps 0;

```
Qy 1 MSKMTSLALLIQLSCYSSSGCGKVLVMPTEFSSHMMNKITLDELIVORGHEVTVLASSA 60
Db 1 MSKMTSLALLIQLSCYSSSGCGKVLVMAAAYSHMMNMKTLLELVORGHEVTVLASSA 60
Qy 61 SISFPNPSPTLKFEVYVSLTKTEFEDIKQLVKRMALPKDTWMSYFSQVQEIIMTFN 120
Db 61 SISFPNPSPTLKFEVYVSLTKTEFEDIKQLVKRMALPKDTWMSYFSQVQEIIMTFN 120
Qy 121 DILKFCDDIVSNKKLMKKLQESRFDVVLADAVPFGGLLAELKIPVYSLRSPGYAI 180
Db 121 DILKFCDDIVSNKKLMKKLQESRFDVVLADAVPFGGLLAELKIPVYSLRSPGYAI 180
Qy 181 EKHSGLLFPSPYVPMVMSLSDDQMTFIERVKMNIYVLYFEFWQIPMKMKDQYSEVL 240
Db 181 EKHSGLLFPSPYVPMVMSLSDDQMTFIERVKMNIYVLYFEFWQIPMKMKDQYSEVL 240
```

```
Qy 181 EKHSGLLFPSPYVPMVMSLSDDQMTFIERVKMNIYVLYFEFWQIPMKMKDQYSEVL 240
Db 181 EKHSGLLFPSPYVPMVMSLSDDQMTFIERVKMNIYVLYFEFWQIPMKMKDQYSEVL 240
Qy 241 GRPPTLSETMAKADIWILIRNWDFOFPHPLPNVEFVGLHCKPAKLPKMEEFVQSSG 300
Db 241 GRPPTLSETMAKADIWILIRNWDFOFPHPLPNVEFVGLHCKPAKLPKMEEFVQSSG 300
Qy 301 ENGVAVPSLGSVMSTSEERANVITASALAKIPQKVLMPDQKPTGLNTRLKYMPION 360
Db 301 ENGVAVPSLGSVMSTSEERANVITASALAKIPQKVLMPDQKPTGLNTRLKYMPION 360
Qy 361 DILGHPKTRAFITTHGNGNGIYEAIYHGVPMVGPPIFGDQDNIAHMKAKGAIVEINFKT 420
Db 361 DILGHPKTRAFITTHGNGNGIYEAIYHGVPMVGPPIFGDQDNIAHMKAKGAIVEINFKT 420
Qy 421 TSEDLRALRTVITDSSYKENAMRLSRIHDDQPKPLDRAVFWIEFWRHKGAGHLSAA 480
Db 421 TSEDLRALRTVITDSSYKENAMRLSRIHDDQPKPLDRAVFWIEFWRHKGAGHLSAA 480
Qy 481 HDLTFQYHSIDVIGFLTCVATAIFLTCKFLFSCCKFNKTRKIEKRE 529
Db 481 HDLTFQYHSIDVIGFLTCVATAIFLTCKFLFSCCKFNKTRKIEKRE 529
```

RESULT 3

```
0969D9 ID 0969D9 PRELIMINARY; PRT; 529 AA.
AC 0969D9;
DT 01-MAR-2001 (TREMBlrel. 16, Created)
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE UDP-glucuronosyltransferase UGT2B33.
OS Macaca mulatta (Rhesus macaque).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
OC Cercopithecinae; Macaca.
CX NCBI_TaxID=9544;
RN [1]
RP SEQUENCE FROM N.A.
RA Dean B., Zhao S., King C.;
RA "Nucleotide and Deduced Amino Acid Sequence of UGT2B33: a Novel UDP-
RT Glucuronosyltransferase Isoform Cloned from Female Rhesus Monkey
RT Liver ";
RL Submitted (AUG-2000) to the EMBL/Genbank/DBJ databases.
DR EMBL; AF294902; AAG21378.1; -.
DR InterPro; IPR002213; UDP_gluco_trans.
DR Pfam; PF00201; UDPGT; 1.
DR PROSITE; PS00375; UDPGT; 1.
KW Transferase.
SQ SEQUENCE 529 AA; 60858 MW; B1956F947F1F78D6 CRC64;
```

Query Match 81.9%; Score 2291; DB 6; Length 529;
Best Local Similarity 81.0%; Pred. No. 4.7e-174;
Matches 427; Conservative 41; Mismatches 57; Indels 2; Gaps 1;

```
Qy 1 MSKMTSLALLIQLSCYSSSGCGKVLVMPTEFSSHMMNKITLDELIVORGHEVTVLASSA 60
Db 1 MSKMTSLALLIQLSCYSSSGCGKVLVMAAAYSHMMNMKTLLELVORGHEVTVLASSA 60
Qy 61 SISFPNPSPTLKFEVYVSLTKTEFEDIKQLVKRMALPKDTWMSYFSQVQEIIMTFN 120
Db 61 SISFPNPSPTLKFEVYVSLTKTEFEDIKQLVKRMALPKDTWMSYFSQVQEIIMTFN 120
Qy 121 DILKFCDDIVSNKKLMKKLQESRFDVVLADAVPFGGLLAELKIPVYSLRSPGYAI 180
Db 121 DILKFCDDIVSNKKLMKKLQESRFDVVLADAVPFGGLLAELKIPVYSLRSPGYAI 180
Qy 181 EKHSGLLFPSPYVPMVMSLSDDQMTFIERVKMNIYVLYFEFWQIPMKMKDQYSEVL 240
Db 181 EKHSGLLFPSPYVPMVMSLSDDQMTFIERVKMNIYVLYFEFWQIPMKMKDQYSEVL 240
```

QY 241 GRPTLSETMAKADIWILIRNWDFOFPHPLPNVEFVGGLHCKPAKPLPKMEEFVQSSG 300
 Db 241 GRHTLSEIMGKADIWILIRNSWNFOFPHPLPNVDFIGGLLCKPAKPLPKMEEFVQSSG 300
 QY 301 ENGIVVSLGMSVNTSEERANVIASALAKIPQKVLWRFDPGNKPTDGLNTRLYKWIPO 360
 Db 301 ENGIVVSLGMSVNTSEERANVIASALAKIPQKVLWRFDPGNKPTDGLNTRLYKWIPO 360
 QY 361 DLGHPKTKAFITHGANGIYEALYHGVPMGVPIFGDQDLNIAHMKAKGAABEINFRTM 420
 Db 361 DLGHPKTKAFITHGANGIYEALYHGVPMGVPIFGDQDLNIAHMKAKGAABEINFRTM 420
 QY 421 TSEDLALARTVIDSSYKENAMRLSRTHDQPVKPLDRAVFWIEFVNRHKGAKHLRGA 480
 Db 421 SSTDANLAKTVINDPLYKENMKLSRIQDQPVKPLDRAVFWIEFVNRHKGAKHLRGA 480
 QY 481 HDLTFQHSIDVIGFLITCVATAIFLTKCFLESCQKFNKTRIEK 527
 Db 481 HDLTFQHSIDVIGFLITCVATAIFLTKCFLESCQKFNKTRIEK 527

RESULT 4

Q9GLE0 PRELIMINARY; PRT; 529 AA.
 ID Q9GLE0
 AC Q9GLE0;
 DT 01-MAR-2001 (TReMBLrel. 16, Created)
 DT 01-MAR-2001 (TReMBLrel. 16, Last sequence update)
 DT 01-MAR-2003 (TReMBLrel. 23, Last annotation update)
 DE UDP-glucuronosyltransferase UGT2B9*2.
 OS Macaca mulatta (Rhesus macaque).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
 OC Cercopithecinae; Macaca.
 NCBI_TaxID=9544;
 RN NCBI [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Liver;
 RA Dean B., Zhao S., King C.;
 RT "Nucleotide and Deduced Amino Acid Sequence of UGT2B9*2: a Novel UDP-
 RT Glucuronosyltransferase Isoform Cloned from Female Rhesus Monkey
 RT Liver.";
 RL Submitted (Aug-2000) to the EMBL/Genbank/DBJ databases.
 DR EMBL; AF294901; AAG21377.1; -
 DR InterPro: IPR002213; UDP_gluco_trans.
 DR Pfam; PF00201; UDPGT; 1.
 DR PROSITE; PS00375; UDPGT; 1.
 DR KMW transferase.
 SQ SEQUENCE 529 AA; 60831 MW; 1225A0163D9AADF2 CRC64;

Query Match 81.8%; Score 2289; DB 6; Length 529;
 Best Local Similarity 80.5%; Pred. No. 6.8e-174; Indels 0; Gaps 0;

Matches 426; Conservative 38; Mismatches 65; Indels 0; Gaps 0;
 QY 1 MSMKWTALLILQISCFYSSGCGKVLVWPFESHMNIKTILDELVORGHEVTVLASSA 60
 Db 1 MSVKMTSYVILLIQISLFYSSGCGKVLVWMAEYHMMNMKTLBELVORGHEVTVLASSA 60
 QY 61 SISDPNSPSTLKEEVYVSLTKTEFEDIKQLVKRWAEPLKDTFWSYFSQVQOIMTFTN 120
 Db 61 SILDPNNSSALKIEVFTSLTKTEFENISQVEVGRWIELPKDTFWLFSQVQOIMTFTN 120
 QY 121 DILKPKCDIVSNKKLKKLOESRFDVVLADAVPFGGLAELLKIPVYSLRSPGYAI 180
 Db 121 DILKPKCDIVSNKKLKKLOESRFDVVLADAVPFGGLAELLKIPVYSLRSPGYAI 180
 QY 181 EKHSGGLFPSPSYVPMVMSLSDQMTFERVKNMIVYLVEFWFOIPFMKKWDQFYSEVL 240
 Db 181 EKHGGGLFPSPSYVPMVMSLSDQMTFERVKNMIVYLVEFWFOIPFMKKWDQFYSEVL 240
 QY 241 GRPTLSETMAKADIWILIRNWDFOFPHPLPNVEFVGGLHCKPAKPLPKMEEFVQSSG 300
 Db 241 GRPTLSETMAKADIWILIRNWDFOFPHPLPNVEFVGGLHCKPAKPLPKMEEFVQSSG 300

QY 301 ENGIVVSLGMSVNTSEERANVIASALAKIPQKVLWRFDPGNKPTDGLNTRLYKWIPO 360
 Db 301 ENGIVVSLGMSVNTSEERANVIASALAKIPQKVLWRFDPGNKPTDGLNTRLYKWIPO 360
 QY 361 DLGHPKTKAFITHGANGIYEALYHGVPMGVPIFGDQDLNIAHMKAKGAABEINFRTM 420
 Db 361 DLGHPKTKAFITHGANGIYEALYHGVPMGVPIFGDQDLNIAHMKAKGAABEINFRTM 420
 QY 421 TSEDLALARTVIDSSYKENAMRLSRTHDQPVKPLDRAVFWIEFVNRHKGAKHLRGA 480
 Db 421 SSTDANLAKTVINDPLYKENMKLSRIQDQPVKPLDRAVFWIEFVNRHKGAKHLRGA 480
 QY 481 HDLTFQHSIDVIGFLITCVATAIFLTKCFLESCQKFNKTRIEK 529
 Db 481 HDLTFQHSIDVIGFLITCVATAIFLTKCFLESCQKFNKTRIEK 529

RESULT 5

Q9TSL6 PRELIMINARY; PRT; 529 AA.
 ID Q9TSL6
 AC Q9TSL6;
 DT 01-MAY-2000 (TReMBLrel. 13, Created)
 DT 01-MAY-2000 (TReMBLrel. 13, Last sequence update)
 DT 01-MAR-2003 (TReMBLrel. 23, Last annotation update)
 DE UDP-glucuronosyltransferase 2B23 (EC 2.4.1.17).
 OS Macaca fascicularis (Crab eating macaque) (Cynomolgus monkey).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
 OC Cercopithecinae; Macaca.
 NCBI_TaxID=9541;
 RN NCBI [1]
 RP SEQUENCE FROM N.A.
 RC MEDLINE=20043918; PubMed=10579317;
 RA Barbier O., Levesque E., Belanger A., Hum D.W.;
 RT "UGT2B23, a novel uridine diphosphate-glucuronosyltransferase enzyme
 RT expressed in steroid target tissues that conjugates androgen and
 RT estrogen metabolites.";
 RL Endocrinology 140:5538-5548 (1999).
 DR EMBL; AF112113; AAT4353.1; -
 DR InterPro: IPR002213; UDP_gluco_trans.
 DR Pfam; PF00201; UDPGT; 1.
 DR PROSITE; PS00375; UDPGT; 1.
 DR KMW Glycosyltransferase; Transferase.
 SQ SEQUENCE 529 AA; 60957 MW; 646315B6D970541A CRC64;

Query Match 80.8%; Score 2261; DB 6; Length 529;
 Best Local Similarity 79.2%; Pred. No. 1.2e-171; Indels 0; Gaps 0;

Matches 419; Conservative 42; Mismatches 68; Indels 0; Gaps 0;

QY 1 MSMKWTALLILQISCFYSSGCGKVLVWPFESHMNIKTILDELVORGHEVTVLASSA 60
 Db 1 MSVKMTSYVILLIQISLFYSSGCGKVLVWMAEYHMMNMKTLBELVORGHEVTVLASSA 60
 QY 61 SISDPNSPSTLKEEVYVSLTKTEFEDIKQLVKRWAEPLKDTFWSYFSQVQOIMTFTN 120
 Db 61 SILDPNNSSALKIEVFTSLTKTEFENISQVEVGRWIELPKDTFWLFSQVQOIMTFTN 120
 QY 121 DILKPKCDIVSNKKLKKLOESRFDVVLADAVPFGGLAELLKIPVYSLRSPGYAI 180
 Db 121 DILKPKCDIVSNKKLKKLOESRFDVVLADAVPFGGLAELLKIPVYSLRSPGYAI 180
 QY 181 EKHSGGLFPSPSYVPMVMSLSDQMTFERVKNMIVYLVEFWFOIPFMKKWDQFYSEVL 240
 Db 181 EKHGGGLFPSPSYVPMVMSLSDQMTFERVKNMIVYLVEFWFOIPFMKKWDQFYSEVL 240
 QY 241 GRPTLSETMAKADIWILIRNWDFOFPHPLPNVEFVGGLHCKPAKPLPKMEEFVQSSG 300
 Db 241 GRHTLSEIMGKADIWILIRNSWNFOFPHPLPNVDFIGGLLCKPAKPLPKMEEFVQSSG 300
 QY 301 ENGIVVSLGMSVNTSEERANVIASALAKIPQKVLWRFDPGNKPTDGLNTRLYKWIPO 360
 Db 301 ENGIVVSLGMSVNTSEERANVIASALAKIPQKVLWRFDPGNKPTDGLNTRLYKWIPO 360

QY 361 DLGHPKTKAFITGNGMGIYEAIYHGVPMVGVPIDFGDQDNDIAHMKAKGAAVEINFKT 420
DB 361 DLGHPKTKAFITGNGMGIYEAIYHGVPMVGIPLEFADQPDNDIAHMKTRGAVALQDFTM 420
QY 421 TSEDLRALRTVITDSYKENAMRLSRHHDOVPKPLDRAVFWIEFWVRHKGAKHLRPA 480
DB 421 SSTDLVNLKTVINDSYKENAMRLSRHHDOVPKPLDRAVFWIEFWVRHKGAKHLRPA 480
QY 481 HDLTFQHSYIDVIGFLITCVATAIFLFTKCFELFSCQKFNKTRKIEKE 529
DB 481 HDLTFQHSYIDVIGFLITCVATAIFLFTKCFELFSCQKFNKTRKIEKE 529

RESULT 6

Q8R084 ID Q8R084 PRELIMINARY; PRT; 529 AA.
AC Q8R084;
DT 01-JUN-2002 (TREMBlrel. 21, Created)
DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE Similar to UDP glycosyltransferase 2 family, polypeptide B17
GN (UDP-glucuronosyltransferase 2B1 precursor).
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RA Strauberg R.;
RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Liver;
RX MEDLINE=22354683; PubMed=12466851;
RA The FANTOM Consortium,
RA the RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs."
RL Nature 420:563-573(2002).
DR EMBL; BC027200; AAH27200.1; -
DR EMBL; AK050435; BAC34254.1; -
DR MGI; MGI:1919023; 1300012D20Rik.
DR InterPro; IPR002213; UDP_gluco_trans.
DR Pfam; PF00201; UDPGT; 1.
DR PROSITE; PS00375; UDPGT; 1.
KW Transferase.
SQ SEQUENCE 529 AA; 60157 MW; B13D713102F794DF CRC64;

Query Match 71.5%; Score 1999.5; DB 11; Length 529;
Best Local Similarity 69.9%; Pred. No. 8.1e-151;
Matches 369; Conservative 62; Mismatches 96; Indels 1; Gaps 1;
QY 1 MSMKMTSA--LLLIQLSCYFSSGCGKVLVWPTFESHMNIKITLDELVOGRHEVTVLASA 60
DB 1 MSMKQASVFLIIQIFCYIRPGACGKVLVWPTFESHMNIKITLDELVOGRHDVTVLISA 60
QY 61 SISDPNPSSTILKEFVYPSLTKTPEFDIIKOLYKRW-ELPKDTFWSYFSQVQEIIMTF 119
DB 61 SILGPSNNESSINFEIYGAPLSKDLEVAFEKVMGNMTYELKKLPFWTSYSLDKRISSEY 120
QY 120 NDILRPFCKDIYSNKKMLKQESRPDVVLADAVPFGELLAELIKIPFVYSLRSPGYA 179
DB 121 SDMKESFCAVAVNWSLKKKQSGKFPDVVLADALVPCBELSELKTVLVSLRCPGYK 180
QY 180 IEKSHSGGLFPSPSYVPMVMSLSDQMTFIERVKMIYVLYEFWFQIFDMKKMDQFYS 239
DB 181 CEKSGGGLPLPSPSYVPMVMSLSDQMTFIERVKMIYVLYEFWFQIFDMKKMDQFYS 240
QY 240 LGRPTTSETMAKADIMLRNYWDFOPHPPLIPNVEFVGLHCKPAKPLPKMEEFVQS 299
DB 241 LGRPTTSETMAKADIMLRNYWDFOPHPPLIPNVEFVGLHCKPAKPLPKMEEFVQS 300

QY 300 GENGVVVSLGSMVSNTESEERANYASALAKIPQKVLMPFGDNKPTDGLNTRLYKMIPO 359
DB 301 GEHGVVVSLSGSMVNIKEBANVVASALAQIPQKVLMPFGDKKPTDGLNTRLYKMIPO 360
QY 360 NDILGHPKTKAFITGNGMGIYEAIYHGVPMVGPIFGDQDNDIAHMKAKGAAVEINFKT 419
DB 361 NDILGHPKTKAFIHAGTNGIYEAIYHGIPIVGIPLEFGDQPDNDINIVAKGAARVDFDT 420
QY 420 MTSDDLRLARTVITDSYKENAMRLSRHHDOVPKPLDRAVFWIEFWVRHKGAKHLRPA 479
DB 421 MSTDDLRLARTVINDSYKENAMRLSRHHDOVPKPLDRAVFWIEFWVRHKGAKHLRPA 480
QY 480 AHDLTWFQHSYIDVIGFLITCVATAIFLFTKCFELFSCQKFNKTRKIEKE 527
DB 481 LHDLTWFQHSYIDVIGFLITCVATAIFLFTKCFELFSCQKFNKTRKIEKE 528

RESULT 7

Q8K154 ID Q8K154 PRELIMINARY; PRT; 532 AA.
AC Q8K154;
DT 01-OCT-2002 (TREMBlrel. 22, Created)
DT 01-OCT-2002 (TREMBlrel. 22, Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE Similar to UDP glycosyltransferase 2 family, polypeptide B4.
GN A1788959.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RA Strauberg R.;
RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC028826; AAH28826.1; -
DR MGI; MGI:2140962; A1788959.
DR InterPro; IPR002213; UDP_gluco_trans.
DR Pfam; PF00201; UDPGT; 1.
DR PROSITE; PS00375; UDPGT; 1.
KW Transferase.
SQ SEQUENCE 532 AA; 60877 MW; 82967F4EEF910A0B CRC64;

Query Match 71.2%; Score 1993.5; DB 11; Length 532;
Best Local Similarity 68.6%; Pred. No. 2.5e-150;
Matches 365; Conservative 78; Mismatches 86; Indels 3; Gaps 2;
QY 1 MSMKMTSA--LLLIQLSCYFSSGCGKVLVWPTFESHMNIKITLDELVOGRHEVTVLAS 58
DB 1 MPVGMTAALLLIQLSGFSGTGKVLVWPMFESHMNIKITLDELKKGHEVTVLRLP 60
QY 59 SASISFPNPSSTILKEFVYPSLTKTPEFDIIKOLYKRW-ELPKDTFWSYFSQVQEIIMTF 117
DB 61 SASISYEVDNLSAIEFTYPSYLSLEBELFWLSLKYATELPKQSFWSGFWFLMQEWMV 120
QY 118 TFNDILRPFCKDIYSNKKMLKQESRPDVVLADAVPFGELLAELIKIPFVYSLRSPG 177
DB 121 VDSKYFESLCKDVAVFNKELMTKQLKSRPDVILADPFIQCDLLAEVLTPLVYSLRFPG 180
QY 178 YAIKSHSGGLFPSPSYVPMVMSLSDQMTFIERVKMIYVLYEFWFQIFDMKKMDQFYS 237
DB 181 STYKYSGGGLPLPSPSYVPMVMSLSDQMTFIERVKMIYVLYEFWFQIFDMKKMDQFYS 240
QY 238 EVLGRPTTSETMAKADIMLRNYWDFOPHPPLIPNVEFVGLHCKPAKPLPKMEEFVQ 297
DB 241 EVLGRPTTSETMAKADIMLRNYWDFOPHPPLIPNVEFVGLHCKPAKPLPKMEEFVQ 300
QY 298 SSGENGVVVSLGSMVSNTESEERANYASALAKIPQKVLMPFGDNKPTDGLNTRLYKMI 357
DB 301 SSGENGVVVSLGSMVSNTESEERANYASALAKIPQKVLMPFGDKKPTDGLNTRLYKMI 360
QY 358 PNDLLGHPKTKAFITGNGMGIYEAIYHGVPMVGPIFGDQDNDIAHMKAKGAAVEINF 417

```

Db      361  PONDLLGSHKTRAFITGSGTNGIYEALHYGIPVAGIPLRQDQYDNIYHLKAKGAARLDF 420
OY      418  KTMISEDLLRALRYITDSYKENAMRLSRIHDDQPVKPLDRAVWIEFWMRHKGAKHLR 477
Db      421  LTMSSTDLTALKTAVNDPSYKENAMRLSRIHDDQPVKPLDRAVWIEFWMRHKGAKHLR 480
OY      478  SAAHLLTMEOHNSIDVIGFLLTCVAATLFTGCFLFSCOKKPKTKIEKE 529
Db      481  VAHHDLSVQYSDVGLFLLACVLTVMFLTKKCCFLCCKKLTGAKKKKG 532

```

| RESULT 8 | | | |
|----------|--|--------------|--------------|
| Q8VIF9 | Q8VIF9 | PRELIMINARY; | PRT; 528 AA. |
| AC | Q8VIF9; | | |
| DT | 01-MAR-2002 (TREMBLrel. 20, Created) | | |
| DT | 01-MAR-2002 (TREMBLrel. 20, last sequence update) | | |
| DT | 01-MAR-2003 (TREMBLrel. 23, last annotation update) | | |
| DE | UDP-glucuronosyltransferase 2B21 (EC 2.7.1.17). | | |
| GN | UGT2B21. | | |
| OS | Cavia porcellus (Guinea pig). | | |
| OC | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; | | |
| OC | Mammalia; Euteria; Rodentia; Hystriognathi; Caviidae; Cavia. | | |
| OX | NCBI_taxid=10141; | | |
| | [1] | | |
| RP | SEQUENCE FROM N.A. | | |
| RC | STRAIN=Hartley; TISSUE=Liver; | | |
| RA | Ishii Y., Miyoshi A., Tsuruda K., Tsuda M., Nagamatsu Y., Yoshisue K., | | |
| RA | Tanaka M., Matsuda S., Ohgaya S., Oguri K.; | | |
| RT | "Hetero-oligomer formation of UDP-glucuronosyltransferases determines | | |
| RT | its substrate specificity: Metabolic activation of morphine to | | |
| RT | morphine-6-glucuronide by UGT2B1 and UGT2B22 hetero-oligomer."; | | |
| RL | submitted (NOV-1999) to the EMBL/GenBank/DBJ databases. | | |
| DR | EMBL; AB034987; BAB82476.1; - | | |
| DR | InterPro; IPR002213; UDP_gluco_trans. | | |
| DR | Pfam; PF00201; UDPGT; 1. | | |
| DR | PROSITE; PS00375; UDPGT; 1. | | |
| KW | Transferase - | | |
| QO | SEQUENCE 528 AA; 60945 MW; 534D7BAAE67ACBB CRC64; | | |

| Query Match | 70.4% | Score 1970.5 | DB 11 | Length 528 |
|-----------------------|-------|--|---------------|------------|
| Best Local Similarity | 68.4% | Pred. No. 1.7e-148 | | |
| Matches | 361 | Conservative 73 | Mismatches 93 | Indels 1 |
| | | | | Gaps 1 |
| Qy | 3 | MKMTSALLLLIOLSCYSSGSCGKVLVMPTEFSHMNNIKITIDELVQGHVTVLASSASI | 62 | |
| Db | 1 | MKRITALLLLLOJCGFHSGSCGKVLVMPMEFHSMMNNIQTILEELIRRGHEVTLRPSCFI | 60 | |
| Qy | 63 | SFDPSPSTLKEPVYPVSLTKTEFPDIIKOLYKRAELPK-DTFMSYSQVQELMTNTD | 121 | |
| Db | 61 | FVDVNTTSEIKFEPTHTFTFRDYEKIFITDLVTLNLTGNSVDTCLDYEPEVEKPKRHS | 120 | |
| Qy | 122 | ILRKPKDIVSNKMKLMKLOESRPFVNLADAVPFGELIAELIKIPFYSLRFSQVAIE | 181 | |
| Db | 121 | EMENVCKELVSNKKMKMLQESRPFILADANGPGELVAELIHTPFYSLRFSQGPQAE | 180 | |
| Qy | 182 | KHSGGLPEPSPVPVYVMSLSDQMTFIERVKMIVLYLFEPMFOQIFDMKMDQFYSEVLG | 241 | |
| Db | 181 | KRAGGLLPSPVPVYVIMSGEMFEMERVKMMICMLYFDFEMFEFDEKRMKLTSEILG | 240 | |
| Qy | 242 | RPTTSETMAKXIDILIRNYMDFOPPHLLPVEVFGVGHCKAPKAPLPKREMEFYQSSGE | 301 | |
| Db | 241 | KPSTILYETMSKXDMMLIRSYMDMEFPHSLPFPDIIIGGLHCKAPKAPLPKREMEFYQSSGE | 300 | |
| Qy | 302 | NGVVVFLSGWVSNSTSEERANVIASALKIPOKVLMPREDGNKPTDILGNTLRLYKVI | 361 | |
| Db | 301 | HGIWVFSIGSMIRNMTDEKANILASALQIPOKVLMPREDGKKPDTILGANTRLRYKVI | 360 | |
| Qy | 362 | LLGHKTKTAFTTHGGMNGIYEIVHGVPMVGPICGQDUNTAHKKAGAAEINFKMT | 421 | |
| Db | 361 | LLGHKTKTAFTTHGGANGIYEIVHGIIPVNGVPLTGPBOYDNTAHKKAGAAKKEFNLS | 420 | |

QY 422 SEDLLRLRLRYITTSSEYKNNMRLSRIRHDDQVPRDLRAVPIIEFMRHKGAGHLSAAH 481

Db 421 STDLLNMAKTYINNFSEYKNNMWLSTIHDDQMRKLDRAVPIIEYVMGKGAHRLPLAH 480

QY 482 DLTFQHSIDVIGFLTLGVATATLFLPKCFPSQCKFKTKTKIEKE 529

Db 481 NLTFQHSIDVIGFLTLGVATATLFLIKCCFLCFQCKFKETGSKKKE 528

| RESULT 9 | | | |
|----------|--|--------------|--------------|
| ID | Q8K169 | PRELIMINARY: | PRT: 530 AA. |
| AC | Q8K169; | | |
| DT | 01-OCT-2002 (TREMBLrel. 22, Created) | | |
| DT | 01-OCT-2002 (TREMBLrel. 22, Last sequence update) | | |
| DT | 01-MAR-2003 (TREMBLrel. 23, Last annotation update) | | |
| DE | UDP-glucuronosyltransferase 2 family, member 5. | | |
| GN | UGT2B5. | | |
| OS | Mus musculus (Mouse). | | |
| OC | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; | | |
| OC | Mammalia; Eutheria; Rodentia; Sclerognathi; Muridae; Murinae; Mus. | | |
| OX | NCBI_TaxId=10090; | | |
| KN | [1] | | |
| RP | SEQUENCE FROM N.A. | | |
| RC | TISSUE=Liver; | | |
| RA | Strausberg R.; | | |
| RL | Submitted (Apr-2002) to the EMBL/GenBank/DBJ databases. | | |
| DR | EMBL; BC028262; AAH28262.1; .. | | |
| DR | MGI; MGI:98900; Ugt2b5. | | |
| DR | InterPro; IPR00213; UDP_gluco_trans. | | |
| DR | Pfam; PF00201; UDPGT; 1. | | |
| DR | PROSITE; PS00375; UDPGT; 1. | | |
| DR | Transferrase. | | |
| SO | SEQUENCE 530 AA; 60874 MW; 99FPA9A8B861D9 CRC64; | | |

| Query Match | 68.2% | Score 1908.5 | DB 11 | Length 530 |
|-----------------------|-----------------|--|----------|------------|
| Best Local Similarity | 67.0% | Pred. No. 1.5e-143 | | |
| Matches 355 | Conservative 66 | Mismatches 108 | Indels 1 | Gaps 1 |
| QY | 1 | MSMKMTSALLLIOLUSCFYSSGSCGKVLVWPTFSHMNNIKTILDELVORGHEVTLASSA | 60 | |
| DB | 1 | MPGKMISALLLILOJISCCRSYVKCGKVLVWPMPEFSHMNNIKTILDELVORGHEVTLARSA | 60 | |
| QY | 61 | SISDPENSPTLTKREYVPVSLTTEBFDIIRQLYKMA-ELPKOTFMSYSQVQEIWTF | 119 | |
| DB | 61 | YVVLDPKPSPLKRETEPPTSVSKNLENFIFIKFVDWVTYEMPRDTCLSYSLNMDIEF | 120 | |
| QY | 120 | NDILRKPCQDIVSKKMLKTIQESRPVVLADAPPCGELAEILKIPVYSLSRPSQYA | 179 | |
| DB | 121 | SDYFLSLCKDVANSKKEMLMTKIQESKFDVLLSDPVASCGELAEILLOIPFLYSIRFSPQYQ | 180 | |
| QY | 180 | IEKSGGLPPPSVYPVPMSELISQOMTIEEKVMIVLYLFEFPMFOIFDMKMPQOFYSEV | 239 | |
| DB | 181 | IEKSGGRFLPPSVYVPIYLSGLGOMTIEIKMMICMLYDFPMQNDKKMSFYSEY | 240 | |
| QY | 240 | LGRPTTISETKAKADWLIRNYWDFOBPBPLLPVNEFVGLHCKPAKPLPKEMEEFYOSS | 299 | |
| DB | 241 | LGRPTTIVETWGQAEWWLIRSNMOLPEPHPLPVNDVYVGLHCKPAKPLPKDMEEFYOSS | 300 | |
| QY | 300 | GENGVVYFSLGSMTSNTSEERANYIASLAKIPQVLMRFPQGNKPDITGLTRLYKTIPO | 359 | |
| DB | 301 | GDHGVVVFSLGSMTSNTTEERANYIAALADIPQVLMKFPQGNKPDITGHTTRVYKMLPO | 360 | |
| QY | 360 | NDLIGHPETKAFIRHGGMNGIYEALITYGVPMVGVPIFGDDOLDNTAHMKAKAAVEINFKT | 419 | |
| DB | 361 | NDLIGHPEPTKAFVIRHGGVYEALITYGIPMTIGPLTGEQDONTAHMVAKAAVAALNIRT | 420 | |
| QY | 420 | MTEEDLLRAITVYITDSSYKENARLSRIHDDQVVKPLDRAVFWIEFVMRHKAQHLRSA | 479 | |
| DB | 421 | MSKSDVNLALBEVIENPFYKKNAMWLSTIHHDQPMKPLDRAVFWIEFVMRHKAQHLRPL | 480 | |
| QY | 480 | AHDLITFOHNSIDVIGLLTCVATALPLFTKCGELFSCQKFNKTRIKRERE | 529 | |

Db 481 GHNLTWQYHSLDVIIGFLSCVATTIVLVCXCLLFIYRFVKENKMKNE 530

RESULT 10

Q91WH2 PRELIMINARY; PRT; 530 AA.

ID Q91WH2 AC Q91WH2;
DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
DE 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE Similar to UDP-glucuronosyltransferase 2 family, member 5.
GN AA986709.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
CX NCBI_TaxID=10090;
NN [1]
RP SEQUENCE FROM N.A.
RC STRAUSBERG R.;
RA Strausberg R.;
RL Submitted (OCT-2001) to the EMBL/Genbank/DBJ databases.
[2]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Embryo;
RX MEDLINE=22354683; PubMed=12466851;
RA The FANTOM Consortium,
RA the RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs."
RL Nature 420:563-573(2002).
DR EMBL; BC015272; AAH15272.1;
DR EMBL; AK034801; BAC28835.1;
DR MGI; MGI:2140794; AA986709.
DR InterPro; IPR002213; UDP_gluco_trans.
DR Pfam; PF00201; UDPGT; 1.
KV Transference.
SQ SEQUENCE 530 AA; 60929 MW; B475642270B28EF1 CRC64;

Query Match 67.7%; Score 1895.5; DB 11; Length 530;
Best Local Similarity 66.8%; Pred. No. 1.6e-142;
Matches 354; Conservative 64; Mismatches 11; Indels 1; Gaps 1;

Qy 1 MSMKWTSAALLIQLSCYFSSGCGKVLWPFESHMNIKITLDELVORGHEVTLAASA 60
Db 1 MPKGISALLLIQISCCFRSVKCGVLPWFMSHMNIKITLDELVORGHEVTLAASA 60

Qy 61 SISFDPNSPSTLKFEVYPVSLTKTEFEDIIKQLVKRA-ELPKDTFWSYFSQVOEIMWT 119
Db 61 YVLDLPKSPGLKFTPTSVSKDNLENFFIKFVDMVTWEMPRDYLCLSYPLQWIDEF 120

Qy 120 NDILAKFCCKDIVSNKKMKKLOESRFDVLAADVPPGELLAEIKIPVYSLRSPGYA 179
Db 121 SDYLSLCKEAVSNKQMTKQESKFDVLSGCELLAEIQLPFLYSIRSPGYQ 180

Qy 180 IEKHSGLLPSPSYVPVMSSELSDQMTFIERVKMNIYLYFEFQIPEMKKMOFYSEV 239
Db 181 IEKSGRFLPSPSYVPVILSGQMTFIERVKMNICHLYDFWQMFNDKMDDFYSEY 240

Qy 240 LGRPTTLETKAKADIMLIRYWDFOFPHPLIPNVEFVGLHCCKPAKPLPKMEEFVQSS 299
Db 241 LGRPTTLETKAKADIMLIRYWDFOFPHPLIPNVEFVGLHCCKPAKPLPKMEEFVQSS 300

Qy 300 GENGIVVPSLGSVMSNTSEERANVIAASAKIPKVLWRFDPGNKDTGLNTRLYKMLPQ 359
Db 301 GDHGVVPSLGSVMSNTSEERANVIAASAKIPKVLWRFDPGNKDTGLNTRLYKMLPQ 360

Qy 360 NDILGHPKTKAFITGGMNGIYEALYHGVPMGVPIFGDQDNIAMKAKGAIVEINPKT 419
Db 361 NDILGHPKTKAFITGGMNGIYEALYHGVPMGVPIFGDQDNIAMKAKGAIVEINPKT 420

Qy 420 MTSDDLRLALRTVITDSSYKENARLSRIHHDQVKKPLDRAVFWIEFVRRHKGAKHLSA 479
Db 421 MTSDDLRLALRTVITDSSYKENARLSRIHHDQVKKPLDRAVFWIEFVRRHKGAKHLSA 480

Qy 480 AHDLTWFOHYSIDVIGFLTCVATAIPLFTKCFELSCQENKFKIEKRE 529
Db 481 GHNLTWQYHSLDVIIGFLSCVATTIVLVCXCLLFIYRFVKENKMKNE 530

RESULT 11

Q98BL9 PRELIMINARY; PRT; 529 AA.

ID Q98BL9 AC Q98BL9;
DT 01-MAR-2003 (TREMBlrel. 23, Created)
DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)
DE 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE Similar to UDP-glucuronosyltransferase 2B5 precursor.
GN Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
CX NCBI_TaxID=10090;
NN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Liver;
RX MEDLINE=22354683; PubMed=12466851;
RA The FANTOM Consortium,
RA the RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs."
RL Nature 420:563-573(2002).
DR EMBL; AK083294; BAC28847.1;
SQ SEQUENCE 529 AA; 60453 MW; 7C246E0716957469 CRC64;

Query Match 67.7%; Score 1895; DB 11; Length 529;
Best Local Similarity 66.4%; Pred. No. 1.7e-142;
Matches 352; Conservative 65; Mismatches 11; Indels 2; Gaps 2;

Qy 1 MSMKWTSAALLIQLSCYFSSGCGKVLWPFESHMNIKITLDELVORGHEVTLAASA 60
Db 1 MPKGISALLLIQISCCFRSVKCGVLPWFMSHMNIKITLDELVORGHEVTLAASA 60

Qy 61 SISFDPNSPSTLKFEVYPVSLTKTEFEDIIKQLVKRA-ELPKDTFWSYFSQVOEIMWT 119
Db 61 YVLDLPKSPGLKFTPTSVSKDNLENFFIKFVDMVTWEMPRDYLCLSYPLQWIDEF 120

Qy 120 NDILAKFCCKDIVSNKKMKKLOESRFDVLAADVPPGELLAEIKIPVYSLRSPGYA 179
Db 121 SDYLSLCKEAVSNKQMTKQESKFDVLSGCELLAEIQLPFLYSIRSPGYQ 179

Qy 180 IEKHSGLLPSPSYVPVMSSELSDQMTFIERVKMNIYLYFEFQIPEMKKMOFYSEV 239
Db 181 IEKSGRFLPSPSYVPVILSGQMTFIERVKMNICHLYDFWQMFNDKMDDFYSEY 239

Qy 240 LGRPTTLETKAKADIMLIRYWDFOFPHPLIPNVEFVGLHCCKPAKPLPKMEEFVQSS 299
Db 241 LGRPTTLETKAKADIMLIRYWDFOFPHPLIPNVEFVGLHCCKPAKPLPKMEEFVQSS 299

Qy 300 GENGIVVPSLGSVMSNTSEERANVIAASAKIPKVLWRFDPGNKDTGLNTRLYKMLPQ 359
Db 301 GDHGVVPSLGSVMSNTSEERANVIAASAKIPKVLWRFDPGNKDTGLNTRLYKMLPQ 359

Qy 360 NDILGHPKTKAFITGGMNGIYEALYHGVPMGVPIFGDQDNIAMKAKGAIVEINPKT 419
Db 361 NDILGHPKTKAFITGGMNGIYEALYHGVPMGVPIFGDQDNIAMKAKGAIVEINPKT 419

Qy 420 MTSDDLRLALRTVITDSSYKENARLSRIHHDQVKKPLDRAVFWIEFVRRHKGAKHLSA 479
Db 421 MTSDDLRLALRTVITDSSYKENARLSRIHHDQVKKPLDRAVFWIEFVRRHKGAKHLSA 479

Qy 480 AHDLTWFOHYSIDVIGFLTCVATAIPLFTKCFELSCQENKFKIEKRE 529
Db 481 AHDLTWFOHYSIDVIGFLTCVATAIPLFTKCFELSCQENKFKIEKRE 529

ID Q8VIF8 PRELIMINARY; PRT; 529 AA.
AC Q8VIF8;
DT 01-MAR-2002 (Tremblrel. 20, Created)
DT 01-MAR-2002 (Tremblrel. 20, Last sequence update)
DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)
DE UDP-glucuronosyltransferase 2B22 (EC 2.7.1.17).
GN UGT2B22.
OS Cavia porcellus (Guinea pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Hystriognathi; Caviidae; Cavia.
OX NCBI_TaxID=10141;
RN [1]
RP SEQUENCE FROM N.A.
RC STFAIN=Hartley; TISSUE=Liver;
RA Ishii Y., Miyoshi A., Tsuruda K., Tenda M., Nagamatsu Y., Yoshitane K.,
Tanaka M., Matsuno S., Ogiya S., Oguri K.,
"hetero-oligomer formation of UDP-glucuronosyltransferases determines
its substrate specificity: Metabolic activation of morphine to
morphine-6-glucuronide by UGT2B21 and UGT2B22 hetero-oligomer";
RL Submitted (NCV-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: AB034988; BAB82477.1; -
DR InterPro: IPR002213; UDP_gluco_trans.
DR Pfam: Pf00201; UDPGT; 1.
DR PROSITE: PS00375; UDPGT; 1.
KM Transferase.
SQ SEQUENCE 529 AA; 60531 MW; 73090A895C16B96 CRC64;

Query Match 67.1%; Score 1878; DB 11; Length 529;
Best Local Similarity 64.5%; Pred. No. 3,9e-141;
Matches 342; Conservative 85; Mismatches 101; Indels 2; Gaps 2;

QY 1 MSNMTSALLLIOLSCYFSSGCGKVLWPTFSHMNNIKITIDELVORGEVTLASSA 60
DB 1 MSNMTSALLLIOLSCYFSSGCGKVLWPTFSHMNNIKITIDELVORGEVTLASSA 60
QY 61 SIFDPSPSTLKEVYVSVLTKEFEDIIKOLVKRMALEPK-DTFWSYFSQVOEIMWTF 119
DB 61 FIFGVNA-SGINEIFPTSIHRDQMEELIKMTKMMIDISIDTCLAVNEMBEILTEY 119
QY 120 NDILRKCKDIYVNSKKMLKQESRPVYLVADAVFPFGELIAELKIPFYSLRFSFGYA 179
DB 120 NDILRKCKDIYVNSKKMLKQESRPVYLVADAVFPFGELIAELKIPFYSLRFSFGYA 179
QY 120 SSVVENCKEIVSNKKLTKLOESKFDVLADPAVPGELVAELIKLPFYVSLRFSFGFQ 179
DB 120 SSVVENCKEIVSNKKLTKLOESKFDVLADPAVPGELVAELIKLPFYVSLRFSFGFQ 179
QY 180 IEKSGGLPPSPYVPMSELSDQMTFERVKMIVLYFEFPOIFDMKKMDQFSEV 239
DB 180 IEKSGGLPPSPYVPMSELSDQMTFERVKMIVLYFEFPOIFDMKKMDQFSEV 239
QY 180 LKRAAGGLPPLPSIIPVTLISLSCQMTFERVKMIVLYFEFPOIFDMKKMDQFSEV 239
DB 180 LKRAAGGLPPLPSIIPVTLISLSCQMTFERVKMIVLYFEFPOIFDMKKMDQFSEV 239
QY 240 LGRPTTSETMAKADIMLRNWDFOFPHPLPNVEFVGLHCKPAKPLPKEMEFPYQS 299
DB 240 LGRPTTSETMAKADIMLRNWDFOFPHPLPNVEFVGLHCKPAKPLPKEMEFPYQS 299
QY 240 LGRPTTSETMAKADIMLRNWDFOFPHPLPNVEFVGLHCKPAKPLPKEMEFPYQS 299
DB 240 LGRPTTSETMAKADIMLRNWDFOFPHPLPNVEFVGLHCKPAKPLPKEMEFPYQS 299
QY 300 GENGUVVFSLSGSMVSNSEEPANVIAALAKIPQKVLMPFGNKPDTGLNTRLYKVIPO 359
DB 300 GENGUVVFSLSGSMVSNSEEPANVIAALAKIPQKVLMPFGNKPDTGLNTRLYKVIPO 359
QY 300 GDHGVVFSLSGSMVSNSEEPANVIAALAKIPQKVLMPFGNKPDTGLNTRLYKVIPO 359
DB 300 GDHGVVFSLSGSMVSNSEEPANVIAALAKIPQKVLMPFGNKPDTGLNTRLYKVIPO 359
QY 360 NDILGHKPTAFITTHGGMNGIYEAIVHGVPMVGPFGDQDNIANHKAKGAIVEINFKT 419
DB 360 NDILGHKPTAFITTHGGMNGIYEAIVHGVPMVGPFGDQDNIANHKAKGAIVEINFKT 419
QY 360 NDILGHKPTAFITTHGGMNGIYEAIVHGVPMVGPFGDQDNIANHKAKGAIVEINFKT 419
DB 360 NDILGHKPTAFITTHGGMNGIYEAIVHGVPMVGPFGDQDNIANHKAKGAIVEINFKT 419
QY 420 MTSDDLRLALRTVITDSYKENAMRLSRHHDPVKPLDRAVFWIEFWMHKGAKHLRSA 479
DB 420 MTSDDLRLALRTVITDSYKENAMRLSRHHDPVKPLDRAVFWIEFWMHKGAKHLRSA 479
QY 420 LSSSDILNALKLVNNPFPYKNALRLSAIHGHQPMKPLDRAVFWIEFWMHKGAKHLRPP 479
DB 420 LSSSDILNALKLVNNPFPYKNALRLSAIHGHQPMKPLDRAVFWIEFWMHKGAKHLRPP 479
QY 480 AHDITWFOHYSIDVIGFLTCVATATFLFTKCLFSPQCKNKRKIEKRE 529
DB 480 AHDITWFOHYSIDVIGFLTCVATATFLFTKCLFSPQCKNKRKIEKRE 529
QY 480 AHDITWFOHYSIDVIGFLTCVATATFLFTKCLFSPQCKNKRKIEKRE 529
DB 480 AHDITWFOHYSIDVIGFLTCVATATFLFTKCLFSPQCKNKRKIEKRE 529

RESULT 13
ID Q8VCN3 PRELIMINARY; PRT; 530 AA.
AC Q8VCN3;
DT 01-MAR-2002 (Tremblrel. 20, Created)

DT 01-MAR-2002 (Tremblrel. 20, Last sequence update)
DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)
DE RIKEN cDNA 0610033E06 gene.
GN 0610033E06RIK.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Kidney;
RA Strausberg R.;
RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: BC019487; AAH19487.1; -
DR MGI: 2148239; 0610033E06RIK.
DR InterPro: IPR002213; UDP_gluco_trans.
DR Pfam: Pf00201; UDPGT; 1.
DR PROSITE: PS00375; UDPGT; 1.
SQ SEQUENCE 530 AA; 60816 MW; A7B8CA04577E3097 CRC64;

Query Match 66.3%; Score 1854.5; DB 11; Length 530;
Best Local Similarity 65.5%; Pred. No. 3e-139;
Matches 347; Conservative 67; Mismatches 115; Indels 1; Gaps 1;

QY 1 MSNMTSALLLIOLSCYFSSGCGKVLWPTFSHMNNIKITIDELVORGEVTLASSA 60
DB 1 MSNMTSALLLIOLSCYFSSGCGKVLWPTFSHMNNIKITIDELVORGEVTLASSA 60
QY 61 SIFDPSPSTLKEVYVSVLTKEFEDIIKOLVKRMA-ELPKDTFWSYFSQVOEIMWTF 119
DB 61 SIFDPSPSTLKEVYVSVLTKEFEDIIKOLVKRMA-ELPKDTFWSYFSQVOEIMWTF 119
QY 61 YVLDPKKSGPLKFEFETPTVTXDDLENFIOPLNVTYELSRDTCISYPLQNMDEL 120
DB 61 YVLDPKKSGPLKFEFETPTVTXDDLENFIOPLNVTYELSRDTCISYPLQNMDEL 120
QY 120 NDILRKCKDIYVNSKKMLKQESRPVYLVADAVFPFGELIAELKIPFYSLRFSFGYA 179
DB 120 NDILRKCKDIYVNSKKMLKQESRPVYLVADAVFPFGELIAELKIPFYSLRFSFGYA 179
QY 121 SGYYLSCKDVVSNRQMLTKLOESKFDVLADPAVPGELIAELKIPFYSLRFSFGYA 180
DB 121 SGYYLSCKDVVSNRQMLTKLOESKFDVLADPAVPGELIAELKIPFYSLRFSFGYA 180
QY 180 IEKSGGLPPSPYVPMSELSDQMTFERVKMIVLYFEFPOIFDMKKMDQFSEV 239
DB 180 IEKSGGLPPSPYVPMSELSDQMTFERVKMIVLYFEFPOIFDMKKMDQFSEV 239
QY 181 IEKSGGRFLPSPYVPMSELSDQMTFERVKMIVLYFEFPOIFDMKKMDQFSEV 240
DB 181 IEKSGGRFLPSPYVPMSELSDQMTFERVKMIVLYFEFPOIFDMKKMDQFSEV 240
QY 240 LGRPTTSETMAKADIMLRNWDFOFPHPLPNVEFVGLHCKPAKPLPKEMEFPYQS 299
DB 240 LGRPTTSETMAKADIMLRNWDFOFPHPLPNVEFVGLHCKPAKPLPKEMEFPYQS 299
QY 241 LGRPTTSETMAKADIMLRNWDFOFPHPLPNVEFVGLHCKPAKPLPKEMEFPYQS 300
DB 241 LGRPTTSETMAKADIMLRNWDFOFPHPLPNVEFVGLHCKPAKPLPKEMEFPYQS 300
QY 300 GENGUVVFSLSGSMVSNSEEPANVIAALAKIPQKVLMPFGNKPDTGLNTRLYKVIPO 359
DB 300 GENGUVVFSLSGSMVSNSEEPANVIAALAKIPQKVLMPFGNKPDTGLNTRLYKVIPO 359
QY 301 GDHGVVFSLSGSMVSNSEEPANVIAALAKIPQKVLMPFGNKPDTGLNTRLYKVIPO 360
DB 301 GDHGVVFSLSGSMVSNSEEPANVIAALAKIPQKVLMPFGNKPDTGLNTRLYKVIPO 360
QY 360 NDILGHKPTAFITTHGGMNGIYEAIVHGVPMVGPFGDQDNIANHKAKGAIVEINFKT 419
DB 360 NDILGHKPTAFITTHGGMNGIYEAIVHGVPMVGPFGDQDNIANHKAKGAIVEINFKT 419
QY 361 NDILGHKPTAFITTHGGMNGIYEAIVHGVPMVGPFGDQDNIANHKAKGAIVEINFKT 420
DB 361 NDILGHKPTAFITTHGGMNGIYEAIVHGVPMVGPFGDQDNIANHKAKGAIVEINFKT 420
QY 420 MTSDDLRLALRTVITDSYKENAMRLSRHHDPVKPLDRAVFWIEFWMHKGAKHLRSA 479
DB 420 MTSDDLRLALRTVITDSYKENAMRLSRHHDPVKPLDRAVFWIEFWMHKGAKHLRSA 479
QY 421 MSRSVDVNLALAEVYENFYPYKNAMVLSITHDQPMKPLDRAVFWIEFWMHKGAKHLRPL 480
DB 421 MSRSVDVNLALAEVYENFYPYKNAMVLSITHDQPMKPLDRAVFWIEFWMHKGAKHLRPL 480
QY 480 AHDITWFOHYSIDVIGFLTCVATATFLFTKCLFSPQCKNKRKIEKRE 529
DB 480 AHDITWFOHYSIDVIGFLTCVATATFLFTKCLFSPQCKNKRKIEKRE 529
QY 481 GHDITWFOHYSIDVIGFLTCVATATFLFTKCLFSPQCKNKRKIEKRE 530
DB 481 GHDITWFOHYSIDVIGFLTCVATATFLFTKCLFSPQCKNKRKIEKRE 530

RESULT 14
ID Q8RI10 PRELIMINARY; PRT; 530 AA.
AC Q8RI10;
DT 01-MAR-2002 (Tremblrel. 13, Created)
DT 01-MAR-2003 (Tremblrel. 13, Last sequence update)
DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)
DE UDP-glucuronosyltransferase UGT2A3.
OS Cavia porcellus (Guinea pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Hystriognathi; Caviidae; Cavia.
OX NCBI_TaxID=10141;

RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Hartley; Tissue=Liver;
 RX MEDLINE=99410697; PubMed=10479484;
 RA Smith S.A., Nagalla S.R., Andrews D.P., Olsen G.D.;
 RT "Morphine regulation of a novel uridine diphosphate glucuronosyl-
 transferase in guinea pig pups following in utero exposure";
 RL Mol. Genet. Metab. 68:68-77(1999).
 DR EMBL; AF175221; AAD5732.1; -;
 DR InterPro; IPR002213; UDP_gluco_trans.
 DR Pfam; PF00201; UDPGT; 1.
 DR PROSITE; PS00375; UDPGT; 1.
 KM Transferase.
 SQ SEQUENCE 530 AA; 59895 MW; EE2F394D3FD484E1 CRC64;
 Query Match 64.4%; Score 1802; DB 11; Length 530;
 Best Local Similarity 65.3%; Pred. No. 4.5e-135;
 Matches 344; Conservative 63; Mismatches 118; Indels 2; Gaps 2;
 QY 4 KMTSLALLIQLSCYSSGCGKLVMPTEFSHNMNIKITLDELVORGEVTVLASSASIS 63
 DB 5 KLASAVLLDLCC-AGSGFCGKLVMPCEMSHMLKTLLELVKRGHEVTVLTLSSNDF 63
 QY 64 FDPNPSPTLKFEVYVSLTKTEFEDIIKQLVKRAE-LPKDTFMSYFSQVQEIWMTFNDI 122
 DB 64 IDYNHRAFNFEVIVPTDKMSSENIINFEI LAVNMPMLQSGKLLQFPVQITED 123
 QY 123 LKRFKQIVSNKKMLKQOESRFDVVLADAVFPFGEELAEELIKTFVYSLRSPGIALEK 182
 DB 124 LGLNRNTVYVQSLMKKLRDSKYDLVLPVPCGELVAEMLVGPVVMLEKPSMGHTTEK 183
 QY 183 HSGGLFPSPYVYVMSSELSDQMTFIERKNNIYVLYFEFVQIDPMKKMDQFYSEVGR 242
 DB 184 YCGQLRAPSPYVPLVGLGILTTMTMERKNNVSEVLDFWLOQDYKFWDFSEALGR 243
 QY 243 PTLSETMAKADIMLRNYWDFQPHPLPNVEFVGLHCKPAKPLPKMEBEFVSSGEN 302
 DB 244 PTLCEINGKAIWIRTYWDFEPRPLPNFEFGHCKPAKPLPKMEBEFVSSGSD 303
 QY 303 GVVVNSLSGMSNTSEBRANVASALAKIPQKYLMPFGNKDPTGLNTRLYKMIPONDL 362
 DB 304 GVVVNSLSGMSNTSEBRANVASALAKIPQKYLMPFGNKDPTGLNTRLYKMIPONDL 363
 QY 363 LGHPTKAFITHGNGNGIYEALYHGVPMVGPVIFGDDLNIAHMKAGAAVEINFKTMS 422
 DB 364 LGHPTKAFITHGNGNGIYEALYHGVPMVGPVIFGDDLNIAHMKAGAAVEINFKTMS 423
 QY 423 EDLRLALRTVITDSSYKKNAMLSRIHNDQVYKPLDRAVFWIEFVNRHKGAGHLSAAD 482
 DB 424 ADLLGALRTVINDPYKKNAMLSRIHNDQVYKPLDRAVFWIEFVNRHKGAGHLSAAD 483
 QY 483 LTFPHYSIDVIGFLLTCVATAIPLFTKCFLESCQKFNKTRKIEKRE 529
 DB 484 LSWFQYHSIDVIGFLLTCVATAIPLFTKCFLESCQKFNKTRKIEKRE 530
 RESULT 15
 Q9Y4X1 PRELIMINARY; PRT; 527 AA.
 AC Q9Y4X1;
 DT 01-NOV-1999 (TRENBLrel. 12, Created)
 DT 01-NOV-1999 (TRENBLrel. 12, Last sequence update)
 DE 01-MAR-2003 (TRENBLrel. 23, Last annotation update)
 DE UDP glucuronosyltransferase (EC 2.4.1.17).
 GN UGT2A1.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
 OX NCB1_Taxid=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC Tissue=Olfactory organ;
 RX MEDLINE=99289328; PubMed=10359671;

RA Jedlicchky G.A., Cassidy A.J., Sales M., Pratt N., Burchell B.;
 RT "Cloning and characterization of a novel human olfactory UDP-
 glucuronosyltransferase";
 RL Biochem. J. 340:837-843(1999).
 DR EMBL; AJ006054; CAB41974.1; -;
 DR Genew; HGNC:12542; UGT2A1.
 DR InterPro; IPR002213; UDP_gluco_trans.
 DR Pfam; PF00201; UDPGT; 1.
 DR PROSITE; PS00375; UDPGT; 1.
 KM Glucosyltransferase; Transferase.
 SQ SEQUENCE 527 AA; 59872 MW; B99E73AA9A421F17 CRC64;
 Query Match 64.2%; Score 1795.5; DB 4; Length 527;
 Best Local Similarity 65.0%; Pred. No. 1.5e-134;
 Matches 330; Conservative 72; Mismatches 103; Indels 3; Gaps 2;
 QY 24 GKVLMPTEFSHNMNIKITLDELVORGEVTVLASSASISFDPNPSPTLKFEVYVSLTK 83
 DB 21 GNVLIWMEGSHMLNVKIIIDELIKKEHNVTVLAVASGALFITPNSNPSLTFEIVKVPFGK 80
 QY 84 TEFEDIIKQLVKRAE-LPKDTFMSYFSQVQEIWMTFNDILRKFKQDIVSNKKMLKQ 141
 DB 81 ERIGVLIKDFVSTVLENRPSPTIWRVYQEMAKVIKDFHVSQETICGVLKNQQLMAK 140
 QY 142 ESREVDVLADAVFPFGEELAEELIKTFVYSLRSPGIALEKHSGLLFPSPYVPMVMSL 201
 DB 141 KSKFEVLSDVPFCGDI VALKLGIPMYSLRFPASVEHGCKVPRPPEVPAVASEL 200
 QY 202 SDQMTFIERKNNIYVLYFEFVQIDPMKKMDQFYSEVGRPTLSETMAKADIMLRNY 261
 DB 201 TDQMSFTDRIRNFISYHLDQYMFETL-WKSDSYYSKALGRPTLCEMTGKAEIWLIRTY 259
 QY 262 WDPQPHPLPNVEFVGLHCKPAKPLPKMEBEFVSSGENGVVFSLSGMSVNTSEBR 321
 DB 260 WDFEPRPYLPNFEFGHCKPAKPLPKMEBEFVSSGENGVVFSLSGMSVNTSEBR 319
 QY 322 NVIASALAKIPQKYLMPFGDKNKPDTGLNTRLYKMIPONDLGHPTKAFITHGNGIY 381
 DB 320 NLIASALAKIPQKYLMPFGDKNKPDTGLNTRLYKMIPONDLGHPTKAFITHGNGIY 379
 QY 382 EAIYHGVPMVGPVIFGDDLNIAHMKAGAAVEINFKTMSSEDLRLARTVITDSSYKEN 441
 DB 380 EAIYHGVPMVGPVIFGDDLNIAHMKAGAAVEINFKTMSSEDLRLARTVITDSSYKEN 439
 QY 442 AMRLSRIHNDQVYKPLDRAVFWIEFVNRHKGAGHLSAADLTFPHYSIDVIGFLLVCV 501
 DB 440 AMRLSRIHNDQVYKPLDRAVFWIEFVNRHKGAGHLSAADLTFPHYSIDVIGFLLVCV 499
 QY 502 ATAIFLFTKCFLESCQKFNKTRKIEKRE 529
 DB 500 TTAIFLVIGCCLFSCQKFGKIGKKKRE 527
 Search completed: December 5, 2003, 09:49:43
 Job time : 41 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 5, 2003, 09:42:55 ; Search time 18 seconds
(without alignments)
1382.063 Million cell updates/sec

Title: US-09-980-729b-5
Perfect score: 2798
Sequence: 1 MSMKMTSALLILQLSCYFSS.....KCFLFSCQKFKTRIEKRE 529

Scoring table: BIOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues
Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Swiseprot_41:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|--------|-------------|--------|----|-------------|
| 1 | 2602.5 | 93.0 | 528 | 1 | UDB4_HUMAN |
| 2 | 2369.5 | 84.7 | 528 | 1 | UDB4_HUMAN |
| 3 | 2330 | 83.3 | 529 | 1 | UDB7_HUMAN |
| 4 | 2297 | 82.1 | 529 | 1 | UDB7_HUMAN |
| 5 | 2292 | 81.9 | 529 | 1 | UDBB_HUMAN |
| 6 | 2281.5 | 81.5 | 529 | 1 | UDBA_HUMAN |
| 7 | 2253 | 80.5 | 529 | 1 | UDBS_HUMAN |
| 8 | 2169.5 | 77.5 | 530 | 1 | UDBH_HUMAN |
| 9 | 2158.5 | 77.1 | 530 | 1 | UDBF_HUMAN |
| 10 | 2100.5 | 75.1 | 530 | 1 | UDBE_HUMAN |
| 11 | 1991.5 | 71.2 | 530 | 1 | UDBE_HUMAN |
| 12 | 1973.5 | 70.5 | 529 | 1 | UDBI_HUMAN |
| 13 | 1938 | 69.3 | 531 | 1 | UDBG_HUMAN |
| 14 | 1938 | 69.3 | 531 | 1 | UDBG_HUMAN |
| 15 | 1930.5 | 69.0 | 530 | 1 | UDBE_HUMAN |
| 16 | 1904.5 | 68.1 | 530 | 1 | UDBS_HUMAN |
| 17 | 1894.5 | 67.7 | 530 | 1 | UDB3_HUMAN |
| 18 | 1856.5 | 66.4 | 530 | 1 | UDB6_HUMAN |
| 19 | 1840.5 | 65.8 | 530 | 1 | UDB6_HUMAN |
| 20 | 1777.5 | 63.0 | 530 | 1 | UDB8_HUMAN |
| 21 | 1761.5 | 63.0 | 527 | 1 | UDB1_HUMAN |
| 22 | 1593 | 56.9 | 502 | 1 | UDB1_HUMAN |
| 23 | 1201.5 | 42.9 | 531 | 1 | UDB1_HUMAN |
| 24 | 1196.5 | 42.8 | 533 | 1 | UDB1_HUMAN |
| 25 | 1188.5 | 41.5 | 533 | 1 | UDB1_HUMAN |
| 26 | 1158 | 41.4 | 535 | 1 | UDB1_HUMAN |
| 27 | 1137.5 | 40.7 | 531 | 1 | UDB1_HUMAN |
| 28 | 1137.5 | 40.7 | 533 | 1 | UDB1_HUMAN |
| 29 | 1135.5 | 40.6 | 535 | 1 | UDB1_HUMAN |
| 30 | 1128 | 40.3 | 534 | 1 | UDB1_HUMAN |
| 31 | 1126 | 40.2 | 534 | 1 | UDB1_HUMAN |
| 32 | 1124 | 40.2 | 534 | 1 | UDB1_HUMAN |
| 33 | 1120.5 | 40.0 | 530 | 1 | UDB1_HUMAN |

| | | | | | | |
|----|--------|------|-----|---|------------|---------------------|
| 34 | 1120.5 | 40.0 | 530 | 1 | UDB1_HUMAN | O6656 homo sapien |
| 35 | 1109.5 | 39.7 | 532 | 1 | UDB1_HUMAN | O28612 oryctolagus |
| 36 | 1099.5 | 39.3 | 530 | 1 | UDB1_HUMAN | O9haw9 homo sapien |
| 37 | 1098.5 | 39.3 | 530 | 1 | UDB1_HUMAN | P36509 homo sapien |
| 38 | 1094.5 | 39.1 | 530 | 1 | UDB1_HUMAN | O9haw9 homo sapien |
| 39 | 1094 | 38.7 | 532 | 1 | UDB1_HUMAN | O62452 mus musculus |
| 40 | 1081.5 | 38.5 | 531 | 1 | UDB1_HUMAN | P19224 homo sapien |
| 41 | 1078.5 | 38.5 | 531 | 1 | UDB1_HUMAN | O64633 rattus norv |
| 42 | 1077.5 | 38.5 | 530 | 1 | UDB1_HUMAN | O64633 rattus norv |
| 43 | 1066 | 38.1 | 531 | 1 | UDB1_HUMAN | O64435 mus musculus |
| 44 | 1045.5 | 37.4 | 529 | 1 | UDB1_HUMAN | P08430 rattus norv |
| 45 | 1045 | 37.3 | 531 | 1 | UDB1_HUMAN | O28611 oryctolagus |

ALIGNMENTS

RESULT 1
UDB4_HUMAN STANDARD; PRT; 528 AA.
AC P06133; O60731; O60867; O75614; P36538;
DT 01-JAN-1988 (Rel. 06, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE UDP-glucuronosyltransferase 2B4 precursor, microsomal (EC 2.4.1.17)
DE (UDP-GT) (Hydroxycholesterol acid) (HLUG25) (UDP-GT-1).
GN UGT2B4 OR UGT2B1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RX MEDLINE=87241362; PubMed=3109396;
RA Jackson M.R., McCarthy L.R., Harding D., Wilson S., Coughtrie M.W.H.,
RT Burchell B.;
RT "Cloning of a human liver microsomal UDP-glucuronosyltransferase
RT cDNA.";
RL Biochem. J. 242:581-586 (1987).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RX MEDLINE=93326164; PubMed=8333863;
RA Jin C.-J., Miners J.O., Lillywhite K.J., McKenzie P.I.;
RT "cDNA cloning and expression of two new members of the human liver
RT UDP-glucuronosyltransferase 2B subfamily.";
RL Biochem. Biophys. Res. Commun. 194:496-503 (1993).
RN [3]
RP SEQUENCE FROM N.A., VARIANT GIU-458, AND CHARACTERIZATION.
RX MEDLINE=99303261; PubMed=10376768;
RA Levesque E., Beaulieu M., Hum D.W., Belanger A.;
RT "Characterization and substrate specificity of UGT2B4 (B458): a
RT UDP-glucuronosyltransferase encoded by a polymorphic gene.";
RL Pharmacogenetics 9:207-216 (1999).
RN [4]
RP SEQUENCE FROM N.A., AND VARIANTS LEU-109 AND LEU-396.
RX McKenzie P.I.;
RA Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.
RN [5]
RP SEQUENCE FROM N.A.
RX Riedy M., Miller A.;
RA "Genomic organization and structure of the UGT2B gene complex at human
RA chromosome 4q13.";
RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
CC -i- FUNCTION: UDP-GTs are of major importance in the conjugation and
CC subsequent elimination of potentially toxic xenobiotics and
CC endogenous compounds. This isozyme is active on polycyclic
CC estrogens (such as estril, 4-hydroxyestrone and 2-hydroxyestril)
CC and xenobiotics (such as 4-methylumbelliferone, 1-naphthol, 4-
CC nitrophenol, 2-aminophenol, 4-hydroxybiphenyl and menthol). It is
CC capable of 6 alpha-hydroxyglucuronidation of hydroxycholesterol.
CC -i- CATALYTIC ACTIVITY: UDP-glucuronate + acceptor = UDP + acceptor

RESULT 2

| | | | | |
|----|--|-----------------------------------|-----------------|------------------------------------|
| ID | UDBJ_MACFA | STANDARD; | PRT; | 528 AA. |
| AC | O9XT55; | | | |
| DT | 30-MAY-2000 | (Rel. 39, Created) | | |
| DT | 30-MAY-2000 | (Rel. 39, Last sequence update) | | |
| DT | 16-OCT-2001 | (Rel. 40, Last annotation update) | | |
| DE | UDP-glucuronosyltransferase 2B19 precursor, microsomal (EC 2.4.1.17). | | | |
| GN | UGT2B19. | | | |
| OS | Carcinoma fascicularis (Crab eating crabapple) (Cynomolgus monkey). | | | |
| OC | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; | | | |
| OC | Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea; | | | |
| OC | Cercopithecinae; Macaca. | | | |
| OK | NCBI_taxid=9541; | | | |
| RN | [1] | | | |
| RP | SEQUENCE FROM N.A. | | | |
| RC | TISSUE=Liver, and Prostate; | | | |
| RX | MEDLINE=99203465; PubMed=10102999; Belanger A.; | | | |
| RA | Belanger G., Barbier O., Hum D.W., | | | |
| RT | "Molecular cloning, expression and characterization of a monkey | | | |
| RT | steroid UDP-glucuronosyltransferase, UGT2B19, that conjugates | | | |
| RU | testosterone."; | | | |
| RL | Eur. J. Biochem. 260:701-708(1999). | | | |
| CC | -I- FUNCTION: UDPGT IS OF MAJOR IMPORTANCE IN THE CONJUGATION AND | | | |
| CC | SUBSEQUENT ELIMINATION OF POTENTIALLY TOXIC XENOBIOTICS AND | | | |
| CC | ENDOGENOUS COMPOUNDS. THIS ISOZYME DISPLAYS ACTIVITY TOWARD | | | |
| CC | SEVERAL CLASSES OF XENOBIOTIC SUBSTRATES: EUGENOL, 4- | | | |
| CC | METHYLMUMBELLIFERONE, P-NITROPHENOL, 1-NAPHTHOL, P'-BIPHENOL, | | | |
| CC | NARINGENIN AND O,O'-BIPHENOL. ACTIVE ALSO ON 3A-HYDROXY AND 17B- | | | |
| CC | HYDROXY POSITIONS OF STEROIDS. | | | |
| CC | -I- FUNCTION: CONTRIBUTES TO THE FORMATION OF ANDROGEN GLUCURONIDE IN | | | |
| CC | EXTRAEPATIC STEROID TARGET TISSUES SUCH AS THE PROSTATE. | | | |
| CC | -I- CATALYTIC ACTIVITY: UDP-glucuronate + acceptor = UDP + acceptor | | | |
| CC | beta-D-glucuronoside. | | | |
| CC | -I- SUBCELLULAR LOCATION: Microsomal (By similarity). | | | |
| CC | -I- TISSUE SPECIFICITY: EXPRESSED IN LIVER, OVARY, PROSTATE, COLON, | | | |
| CC | KIDNEY, PANCREAS, BRAIN, CEREBELLUM, MAMMARY GLAND AND EPIDIDYMIS. | | | |
| CC | NOT EXPRESSED IN SMALL INTESTINE, SPLEEN, BLADDER, ADRENAL GLAND | | | |
| CC | AND TESTIS. | | | |
| CC | -I- SIMILARITY: BELONGS TO THE UDP-GLYCOSYLTRANSFERASE FAMILY. | | | |
| CC | ----- | | | |
| CC | This SWISS-PROT entry is copyright. It is produced through a collaboration | | | |
| CC | between the Swiss Institute of Bioinformatics and the EMBL Outstation - | | | |
| CC | the European Bioinformatics Institute. There are no restrictions on its | | | |
| CC | use by non-profit institutions as long as its content is in no way | | | |
| CC | modified and this statement is not removed. Usage by and for commercial | | | |
| CC | entities requires a license agreement (See http://www.isb-sib.ch/announce/ | | | |
| CC | or send an email to license@isb-sib.ch). | | | |
| CC | ----- | | | |
| DR | EMBL; AF112112; AAD24435.1; - | | | |
| DR | InterPro; IPR002213; UDP_Gluco_trans. | | | |
| DR | Pfam; PF00201; UDPGT; 1. | | | |
| DR | PROSITE; PS00375; UDPGT; 1. | | | |
| KW | Transferase; Glycosyltransferase; Glycoprotein; Transmembrane; Signal; | | | |
| KW | MultiGene family; Microsome. | | | |
| FT | SIGNAL | 1 | 21 | POTENTIAL. |
| FT | CHAIN | 22 | 528 | UDP-GLUCURONOSYLTRANSFERASE 2B19. |
| FT | TRANSMM | 493 | 513 | POTENTIAL (GLCNAC...) (POTENTIAL). |
| FT | CARBOHD | 315 | 315 | N-LINKED (GLCNAC...) (POTENTIAL). |
| SQ | SEQUENCE | 528 AA; | 3BFDAE714A27ABE | CKG64; |

Query Match 84.7%; Score 2369.5; DB 1; Length 528;
 Best Local Similarity 82.8%; Pred. No. 7.5e-112;
 Matches 438; Conservative 43; Mismatches 47; Indels 1; Gaps 1;

QY 1 MSKMTSALLLIQSCYSSGCGKVLVWPTFSHMNNIKITILDELVOGHEVTVLASSA 60
 DB 1 MSKMTSALLLIQSCYSSGCGKVLVWPTFSHMNNIKITILDELVOGHEVTVLASSA 60
 QY 61 SISFDPSPSTLAKFEVYVSLTKTEFEDIIKOLVKRAELPKDPMFVSFQVOGIMMTFN 120
 DB 61 SISFDPSPSTLAKFEVYVSLTKTEFEDIIKOLVKRAELPKDPMFVSFQVOGIMMTFN 120
 QY 121 DILKPKCDIVSNKLMKLOESRFDVVLADAVFPFGLAELLIKIPVYSLRSPGYAI 180
 DB 121 DILKPKCDIVSNKLMKLOESRFDVVLADAVFPFGLAELLIKIPVYSLRSPGYAI 180
 QY 181 EKHSGLLPPSPYVPMVMSLSDQMTFIERVKMNIYVLYEFWFQIFDMKKMDQFYSEVL 240
 DB 181 EKHSGLLPPSPYVPMVMSLSDQMTFIERVKMNIYVLYEFWFQIFDMKKMDQFYSEVL 240
 QY 241 GRPTLSETMAKADIVLIRNWDQFPHPLPNVEFVGLHCKRAKPLPKMEEFVVOSSG 300
 DB 241 GRPTLSETMAKADIVLIRNWDQFPHPLPNVEFVGLHCKRAKPLPKMEEFVVOSSG 300
 QY 301 ENGVSFSLGSMVNTSEERANVIASALAKIPQKVLRFPGDKPDTGLNTRLYKWIIPON 360
 DB 301 ENGVSFSLGSMVNTSEERANVIASALAKIPQKVLRFPGDKPDTGLNTRLYKWIIPON 360
 QY 361 DLGHPKTKAFITHGNGIYEALYHGVPMVGPVIFGDQDNLNIAHMKAKGAAVEINFKTM 420
 DB 361 DLGHPKTKAFITHGNGIYEALYHGVPMVGPVIFGDQDNLNIAHMKAKGAAVEINFKTM 420
 QY 421 TSEDLALRVTIDSSYKENAMRLSRHHDQPKYKPLDRANFVLEFVNRHKGAGHLASAA 480
 DB 421 TSEDLALRVTIDSSYKENAMRLSRHHDQPKYKPLDRANFVLEFVNRHKGAGHLASAA 480
 QY 481 HDLTFPOHYSIDVAGLLTCVATAIPLFTKCFLLFSCCKENRTKIEKRE 529
 DB 481 HDLTFPOHYSIDVAGLLTCVATAIPLFTKCFLLFSCCKENRTKIEKRE 529
 DB 481 HDLTFPOHYSIDVAGLLTCVATAIPLFTKCFLLFSCCKENRTKIEKRE 529

RESULT 3
 UDB7 HUMAN
 ID UDB7 HUMAN STANDARD: PRT; 529 AA.
 AC P16662;
 DT 01-AUG-1990 (Rel. 15, Created)
 DT 01-AUG-1990 (Rel. 15, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE UDP-glucuronosyltransferase 2B7 precursor, microsomal (EC 2.4.1.17)
 DE (UDPGT) (3,4-catechol estrogen specific) (UDPGTH-2).
 GN UGT2B7 OR UGT1B2B9.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 OC NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Liver;
 RX MEDLINE=90243659; PubMed=2159463;
 RA Raliter J.K., Sheen Y.Y., Owens I.S.;
 RT "Cloning and expression of human liver UDP-glucuronosyltransferase in
 RT COS-1 cells. 3,4-catechol estrogens and estradiol as primary
 RT substrates.";
 RL J. Biol. Chem. 265:7900-7906(1990).
 RN [2]
 RP VARIANT UGT2B7*2.
 RX MEDLINE=21033487; PubMed=11186130;
 RA Bhasker C.R., McKinnon W., Stone A., Lo A.C., Kubota T., Ishizaki T.,
 RA Miners J.O.;
 RT "Genetic polymorphism of UDP-glucuronosyltransferase 2B7 (UGT2B7) at
 RT amino acid 268: ethnic diversity of alleles and potential clinical
 RT significance.";

RL Pharmacogenetics 10:679-685(2000).
 CC -1- FUNCTION: UDPGT IS OF MAJOR IMPORTANCE IN THE CONJUGATION AND
 CC SUBSEQUENT ELIMINATION OF POTENTIALLY TOXIC XENOBIOTICS AND
 CC ENDOGENOUS COMPOUNDS.
 CC -1- FUNCTION: ITS UNIQUE SPECIFICITY FOR 3,4-CATECHOL ESTROGENS AND
 CC ESTRIOL SUGGESTS IT MAY PLAY AN IMPORTANT ROLE IN REGULATING THE
 CC LEVEL AND ACTIVITY OF THESE POTENT AND ACTIVE ESTROGEN
 CC METABOLITES.
 CC -1- CATALYTIC ACTIVITY: UDP-glucuronate + acceptor = UDP + acceptor
 CC -1- SUBCELLULAR LOCATION: Microsomal.
 CC -1- SIMILARITY: BELONGS TO THE UDP-GLYCOSYLTRANSFERASE FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL, J05428; AAA36793.1; -.
 DR PIR, A35366; A35366.
 DR Genew; HGNC:12554; UGT2B7.
 DR MIM; 600068; -.
 DR GO; GO:0005624; C:membrane fraction; TAS.
 DR GO; GO:0006629; P:lipid metabolism; TAS.
 DR InterPro; IPR002213; UDP_gluco_trans.
 DR Pfam; PF00201; UDPGT; 1.
 DR PROSITE; PS00375; UDPGT; 1.
 DR KX Transferrase; Glycosyltransferase; Glycoprotein; Transmembrane; Signal;
 DR Multigene family; Microsome; Polymorphism.
 FT CHAIN 1 23
 FT TRANSEM 24 529
 FT CARBOHYD 67 67
 FT CARBOHYD 68 68
 FT CARBOHYD 315 315
 FT VARIANT 268 268
 FT H -> Y (IN ALLELE UGT2B7*2).
 FT FTID=VAR_012342.
 SQ SEQUENCE 529 AA; 60694 MW; 94B1CA6CE92F1446 CRC64;

Query Match 83.3%; Score 2330; DB 1; Length 529;
 Best Local Similarity 81.1%; Pred. No. 7.4e-169;
 Matches 429; Conservative 42; Mismatches 58; Indels 0; Gaps 0;

QY 1 MSKMTSALLLIQSCYSSGCGKVLVWPTFSHMNNIKITILDELVOGHEVTVLASSA 60
 DB 1 MSKMTSALLLIQSCYSSGCGKVLVWPTFSHMNNIKITILDELVOGHEVTVLASSA 60
 QY 61 SISFDPSPSTLAKFEVYVSLTKTEFEDIIKOLVKRAELPKDPMFVSFQVOGIMMTFN 120
 DB 61 SISFDPSPSTLAKFEVYVSLTKTEFEDIIKOLVKRAELPKDPMFVSFQVOGIMMTFN 120
 QY 121 DILKPKCDIVSNKLMKLOESRFDVVLADAVFPFGLAELLIKIPVYSLRSPGYAI 180
 DB 121 DILKPKCDIVSNKLMKLOESRFDVVLADAVFPFGLAELLIKIPVYSLRSPGYAI 180
 QY 181 EKHSGLLPPSPYVPMVMSLSDQMTFIERVKMNIYVLYEFWFQIFDMKKMDQFYSEVL 240
 DB 181 EKHSGLLPPSPYVPMVMSLSDQMTFIERVKMNIYVLYEFWFQIFDMKKMDQFYSEVL 240
 QY 241 GRPTLSETMAKADIVLIRNWDQFPHPLPNVEFVGLHCKRAKPLPKMEEFVVOSSG 300
 DB 241 GRPTLSETMAKADIVLIRNWDQFPHPLPNVEFVGLHCKRAKPLPKMEEFVVOSSG 300
 QY 301 ENGVSFSLGSMVNTSEERANVIASALAKIPQKVLRFPGDKPDTGLNTRLYKWIIPON 360
 DB 301 ENGVSFSLGSMVNTSEERANVIASALAKIPQKVLRFPGDKPDTGLNTRLYKWIIPON 360
 QY 361 DLGHPKTKAFITHGNGIYEALYHGVPMVGPVIFGDQDNLNIAHMKAKGAAVEINFKTM 420
 DB 361 DLGHPKTKAFITHGNGIYEALYHGVPMVGPVIFGDQDNLNIAHMKAKGAAVEINFKTM 420

```

OY 61 S1SDPNPSPLTKREYV PVLSTKTEEPDII KOLYKRAAEI PKOTFMSYFSSOVGSIIMTPTN 120
Db 61 SILFDPNNSSALKI EIEPPTSLTKTEFENISMQEYKRI ELPKDTFWLYFSQMOEIMRFG 120
OY 121 DILRFCKDVIYSNKKLKKLOESRFDVLADAVPFGEELAEELIKIEFVYSLRSPGYAI 180
Db 121 DIINFCDDVSNKKLKKLOESRFDVADPIRFGSELLAEELFNILVYSLRFTPGYIF 180
OY 181 EKHSGLLIPPSYVPVVMSELSDMTIERVKMIYVLVEEFWFQIDMKKWDQFYSEVL 240
Db 181 EKHGGELFPSPYVPVVMSELSDMTFEMERKNNIYMLSPDFYQWYDMKKMDQFYSEVL 240
OY 241 GRPPTLSEYMAKAIWILIRNWDQFPHPLIPNVEFVGLHCKRAKPLPKMESEFVSSG 300
Db 241 GRPPTLSEYMAKAIWILIRNWNQFPHPLIPNVEFVGLHCKRAKPLPKMESEFVSSG 300
OY 301 ENGVVVSISGMVNTSEERANVASALAKIPQKYLRFDSNKPDTJGLNRLYKWIIPON 360
Db 301 ENGVVVSISGMVNTSEERANVASALAKIPQKYLRFDSNKPDTJGLNRLYKWIIPON 360
OY 361 DLIGHPTKATITGGMNGIYEALYHGVPMVGPIDGDDLNIAHMKAGAAVEINFKTM 420
Db 361 DLIGHPTKATITGGMNGIYEALYHGVPMVGPIDGDDLNIAHMKAGAAVEINFKTM 420
OY 421 TSEDLRLARVITDDSYKENAMLSIIHHQPKPLDRARVFWLEFVMBRHGAHLSAA 480
Db 421 TSEDLRLARVITDDSYKENAMLSIIHHQPKPLDRARVFWLEFVMBRHGAHLSAA 480
OY 481 HDLTFQHSIDVIGFLLTCVATAIFPTKCFELSCOKFNKTRIEKRE 529
Db 481 HDLTFQHSIDVIGFLLTCVATAIFPTKCFELSCOKFNKTRIEKRE 529
OY 481 HDLTFQHSIDVIGFLLTCVATAIFPTKCFELSCOKFNKTRIEKRE 529
Db 481 HDLTFQHSIDVIGFLLTCVATAIFPTKCFELSCOKFNKTRIEKRE 529

RESULT 5
UBBB_HUMAN
ID _UBBB_HUMAN STANDARD: PRT: 529 AA.
AC 075310.
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE UDP-glucuronosyltransferase 2B11 precursor, microsomal (EC 2.4.1.17)
EN (UDPGT).
GN UGT2B11.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Cateniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OC NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=98340847; Pubmed=9675083;
RA Beaulieu M., Levesque E., Hum D.W., Belanger A.;
RT "Isolation and characterization of a human orphan UDP-
RT glucuronosyltransferase, UGT2B11."
RL Biochem. Biophys. Res. Commun. 248:44-50(1998).
CC -!- FUNCTION: UDPGT IS OF MAJOR IMPORTANCE IN THE CONJUGATION AND
CC SUBSEQUENT ELIMINATION OF POTENTIALLY TOXIC XENOBIOTICS AND
CC ENDOGENOUS COMPOUNDS.
CC -!- CATALYTIC ACTIVITY: UDP-glucuronate + acceptor = UDP + acceptor
CC beta-D-glucuronoside.
CC -!- SUBCELLULAR LOCATION: Microsomal.
CC -!- TISSUE SPECIFICITY: Widely expressed.
CC -!- SIMILARITY: BELONGS TO THE UDP-GLYCOSYLTRANSFERASE FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation at
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb.ch/announce/
CC or send an email to license@isb.ch).
CC EMBL; AF016492; AAC27891.1; --
DR Genew; HGNC:12545; UGT2B11.

```

DR MIM; 603064; -.
 DR GO; GO:0008210; P:estrogen metabolism; TAS.
 DR GO; GO:0006805; P:xenobiotic metabolism; TAS.
 DR InterPro; IPR002213; UDP_gluco_trans.
 DR Pfam; PF00201; UDPGT; 1.
 DR PROSITE; PS00375; UDPGT; 1.
 DR Transferrase; Glycosyltransferase; Glycoprotein; Transmembrane; Signal;
 KM Multigene family; Microsome.
 FT SIGNAL 1 21 POTENTIAL.
 FT CHAIN 1 21 UDP-GLUCURONOSYLTRANSFERASE 2B11.
 FT TRANSMEM 493 513 POTENTIAL.
 FT CARBOHYD 315 315 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 529 AA; 61038 MW; CE4AC3C71CFC2AB4 CRC64;
 Query Match 81.9%; Score 2292; DB 1; Length 529;
 Best Local Similarity 80.2%; Pred. No. 5.6e-165;
 Matches 424; Conservative 43; Mismatches 62; Indels 0; Gaps 0;
 QY 1 MSMKMTSALLLIOLSCYFSSGCGKVLVWPTFESHMNNKITIIDELOYRGHEVTVLASSA 60
 DB 1 MTKMTSVLLIHLISCFYSSGCGKVLVWMAEYSHMNMKITILKELOYRGHEVTVLASSA 60
 QY 61 SISDPNPSSTLKEVYVSVLTKTEFEDIYQLYKNAELPKDTFWSYFSQVQSEIMWTFN 120
 DB 61 SILFDPNDASTLKEVYVPTSLKTEFENIMQYKRWSDIRKDSFWLTFYFQSEIIMELY 120
 QY 121 DILKPFCDIVSNKKLMKKLOESRFDVLAFAVPPGELLALNIRVYSLRFTPGYTI 180
 DB 121 DIFNPFCDIVSNKKLMKKLOESRFDVLAFAVPPGELLALNIRVYSLRFTPGYTI 180
 QY 181 EKHSGLLFPSPYVYVVMSELSDMQTFIERVKNNIYVLYFEFPMQIDPMKKDQFYSEVL 240
 DB 181 ERHSGGLFPSPYPIVYVMSKLSDMQTFIERVKNNIYVLYFEFPMQIDPMKKDQFYSEVL 240
 QY 241 GRPPTLSETMAKADIMLRNWDQFPHPLPNVEFVGGLHCKPAKLPKMEBEFVQSSG 300
 DB 241 GRPPTLSETMAKADIMLRNWDQFPHPLPNVEFVGGLHCKPAKLPKMEBEFVQSSG 300
 QY 241 GRPPTLSETMAKADIMLRNWDQFPHPLPNVEFVGGLHCKPAKLPKMEBEFVQSSG 300
 DB 301 ENGVAVSIGSMVNTSEERANVIASALAKIPQVLMRFPDGNKPTDGLNTRLYKWIIPON 360
 DB 301 ENGVAVSIGSMVNTSEERANVIASALAKIPQVLMRFPDGNKPTDGLNTRLYKWIIPON 360
 QY 361 DLLGHPKTKAFITHGGMNGIYEALYHGVPMGVPIFGDQDNIAMHAKAGAIVEINFKTM 420
 DB 361 DLLGHPKTKAFITHGGMNGIYEALYHGVPMGVPIFGDQDNIAMHAKAGAIVEINFKTM 420
 QY 421 TSDDLRLALRTVITDSSYKENAMRLSRIHHDQPVKPLDRAVFWIEFVWRHKGAGHLSSAA 480
 DB 421 SSTDLRLALRTVINDPSYKENIMKLSRIHQDQPVKPLDRAVFWIEFVWRHKGAGHLSSAA 480
 QY 481 HDLTFWFGHYSTIDVIGLLTCAVTAIFLTKCFLSCQKFNKTRKLEKRE 529
 DB 481 HDLTFWFGHYSTIDVIGLLTCAVTAIFLTKCFLSCQKFNKTRKLEKRE 529
 RESULT 6
 ID UDBA_HUMAN STANDARD; PRT; 528 AA.
 AC P36537;
 DT 01-JUN-1994 (Rel. 29, Created)
 DT 01-JUN-1994 (Rel. 29, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE UDP-glucuronosyltransferase 2B10 precursor, microsomal (EC 2.4.1.17)
 DE (UDPGT).
 GN UGT2B10.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Liver;
 RX MEDLINE=93326164; PubMed=8333863;

RA Jin C.-J., Miners J.O., Lillywhite K.J., McKenzie P.I.;
 RT "cDNA cloning and expression of two new members of the human liver
 RT UDP-glucuronosyltransferase 2B subfamily".
 RL Biochem. Biophys. Res. Commun. 194:496-503(1993).
 CC -1- FUNCTION: UDPGT IS OF MAJOR IMPORTANCE IN THE CONJUGATION AND
 CC SUBSEQUENT ELIMINATION OF POTENTIALLY TOXIC XENOBIOTICS AND
 CC ENDOGENOUS COMPOUNDS.
 CC -1- CATALYTIC ACTIVITY: UDP-glucuronate + acceptor = UDP + acceptor
 CC beta-D-glucuronoside.
 CC -1- SUBCELLULAR LOCATION: Microsomal.
 CC -1- SIMILARITY: BELONGS TO THE UDP-GLYCOSYLTRANSFERASE FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; X63359; CAA44961.1; -.
 DR PIR; JN0620; JN0620.
 DR Genew; HGNC:12544; UGT2B10.
 DR MIM; 600070; -.
 DR GO; GO:0006629; P:liver metabolism; TAS.
 DR InterPro; IPR002213; UDP_gluco_trans.
 DR Pfam; PF00201; UDPGT; 1.
 DR PROSITE; PS00375; UDPGT; 1.
 DR Transferrase; Glycosyltransferase; Glycoprotein; Transmembrane; Signal;
 KM Multigene family; Microsome.
 FT SIGNAL 1 23 BY SIMILARITY.
 FT CHAIN 24 528 UDP-GLUCURONOSYLTRANSFERASE 2B10.
 FT TRANSMEM 492 512 POTENTIAL.
 FT CARBOHYD 66 66 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 314 314 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 481 481 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 528 AA; 60774 MW; 5B6EA3DBC032C2B0 CRC64;
 Query Match 81.5%; Score 2281.5; DB 1; Length 528;
 Best Local Similarity 80.5%; Pred. No. 3.3e-165;
 Matches 426; Conservative 44; Mismatches 58; Indels 1; Gaps 1;
 QY 1 MSMKMTSALLLIOLSCYFSSGCGKVLVWPTFESHMNNKITIIDELOYRGHEVTVLASSA 60
 DB 1 MALKMTT-VLLIOLSFYSSGCGKVLVWMAEYSHMNMKITILKELOYRGHEVTVLASSA 59
 QY 61 SISDPNPSSTLKEVYVSVLTKTEFEDIYQLYKNAELPKDTFWSYFSQVQSEIMWTFN 120
 DB 60 SILFDPNDASTLKEVYVPTSLKTEFENIMQYKRWSDIRKDSFWLTFYFQSEIIMELY 119
 QY 121 DILKPFCDIVSNKKLMKKLOESRFDVLAFAVPPGELLALNIRVYSLRFTPGYTI 180
 DB 120 DIFNPFCDIVSNKKLMKKLOESRFDVLAFAVPPGELLALNIRVYSLRFTPGYTI 179
 QY 181 EKHSGLLFPSPYVYVVMSELSDMQTFIERVKNNIYVLYFEFPMQIDPMKKDQFYSEVL 240
 DB 180 ERHSGGLFPSPYPIVYVMSKLSDMQTFIERVKNNIYVLYFEFPMQIDPMKKDQFYSEVL 239
 QY 241 GRPPTLSETMAKADIMLRNWDQFPHPLPNVEFVGGLHCKPAKLPKMEBEFVQSSG 300
 DB 240 GRPPTLSETMAKADIMLRNWDQFPHPLPNVEFVGGLHCKPAKLPKMEBEFVQSSG 299
 QY 301 ENGVAVSIGSMVNTSEERANVIASALAKIPQVLMRFPDGNKPTDGLNTRLYKWIIPON 360
 DB 300 ENGVAVSIGSMVNTSEERANVIASALAKIPQVLMRFPDGNKPTDGLNTRLYKWIIPON 359
 QY 361 DLLGHPKTKAFITHGGMNGIYEALYHGVPMGVPIFGDQDNIAMHAKAGAIVEINFKTM 420
 DB 360 DLLGHPKTKAFITHGGMNGIYEALYHGVPMGVPIFGDQDNIAMHAKAGAIVEINFKTM 419
 QY 421 TSDDLRLALRTVITDSSYKENAMRLSRIHHDQPVKPLDRAVFWIEFVWRHKGAGHLSSAA 480
 DB 420 SSTDLRLALRTVINDPSYKENIMKLSRIHQDQPVKPLDRAVFWIEFVWRHKGAGHLSSAA 479

RA Hum D.W.;
 RT "Chromosomal localization, structure, and regulation of the UGT2B17
 RT gene, encoding a C19 steroid metabolizing enzyme.";
 RL DNA Cell Biol. 16:1143-1154(1997).
 CC -1- FUNCTION: UDPGT IS OF MAJOR IMPORTANCE IN THE CONJUGATION AND
 CC SUBSEQUENT ELIMINATION OF POTENTIALLY TOXIC XENOBIOTICS AND
 CC ENDOGENOUS COMPOUNDS. THE MAJOR SUBSTRATES OF THIS ISOZYME ARE
 CC EUGENOL > 4-METHYLBELLEREPONE > DIHYDROTESTOSTERONE (DHT) >
 CC ANDROSTANE-3ALPHA,17BETA-DIOL (3ALPHA-DIOL) > TESTOSTERONE >
 CC ANDROSTERONE (ADT).
 CC CATALYTIC ACTIVITY: UDP-glucuronate + acceptor = UDP + acceptor
 CC beta-D-glucuronoside.
 CC -1- SUBCELLULAR LOCATION: Microsomal.
 CC TISSUE SPECIFICITY: EXPRESSED IN VARIOUS TISSUES INCLUDING THE
 CC LIVER, KIDNEY, TESTIS, UTERUS, PLACENTA, MAMMARY GLAND, ADRENAL
 CC GLAND, SKIN, AND PROSTATE.
 CC -1- SIMILARITY: BELONGS TO THE UDP-GLYCOSYLTRANSFERASE FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL collaboration
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: U59209; AAC25491.1; -
 CC Genew; HGNC:12547; UGT2B17.
 CC DR MIM: 601903; -
 CC DR GO: GO:0005624; C:membrane fraction; TAS.
 CC DR GO: GO:0008202; P:steroid metabolism; TAS.
 CC DR InterPro: IPR002213; UDP_gluco_trans.
 CC DR Pfam: PF00201; UDPGT; 1.
 CC DR PROSITE: PS00375; UDPGT; 1.
 CC KM Transferase; Glycosyltransferase; Glycoprotein; Transmembrane; Signal;
 CC MultiGene family; Microsome.
 CC FT SIGNAL 1 23 POTENTIAL.
 CC FT CHAIN 1 24 530 UDP-GLUCURONOSYLTRANSFERASE 2B17.
 CC FT TRANSMEM 495 515 POTENTIAL.
 CC FT CARBOHYD 65 65 N-LINKED (GLCNAC. . .) (POTENTIAL).
 CC FT CARBOHYD 316 316 N-LINKED (GLCNAC. . .) (POTENTIAL).
 CC FT CARBOHYD 483 483 N-LINKED (GLCNAC. . .) (POTENTIAL).
 CC SQ SEQUENCE 530 AA; 61095 MW; 8E59EBCA43FA3760 CRC64;
 CC
 CC Query Match 77.5%; Score 2169.5; DB 1; Length 530;
 CC Best Local Similarity 74.7%; Pred. No. 11e-156;
 CC Matches 396; Conservative 60; Mismatches 73; Indels 1; Gaps 1;
 CC
 CC QY 1 MSMKWTSLALLIOLSCYSSGCGVLVWPFESHMNIKTITLDELVORGHVTVLASSA 60
 CC Db 1 MSLLKMSVFLMLQSLSCFSSGCGVLVWPFESHMNIKTITLDELVORGHVTVLASSA 60
 CC 61 SIFSDPNSPTLKEVYVPSLTKEFEEDIIKOLYKRA-ELPKDTFMSYFSQVOEIMWT 119
 CC Db 61 SILVNASKSAIKLEVPYPTSLTKNDLDEPFMKFMDRWYTSISKDTFMSYFQDLQCEY 120
 CC QY 120 NDILRKCKDIYVSKKMLKQESRPDVVLDANVPPELLAEILKIPFVTSLRSPRYA 179
 CC Db 121 SDVYIKCEDAVLNKKMRKIQESKFDVLDANVPCELLAEILNIPFLVSLRFSVGYT 180
 CC 180 IEKSSGGLPPSPVYVWSELSDOMTIERVKNNIYLYPEPFQIPDMKKMOQFSEV 239
 CC Db 181 VERKGGGLPPSPVYVWSELSDOMTIERVKNNIYLYPEPFQIPDMKKMOQFSEV 240
 CC QY 240 LGRPTTSETMAKADIMLRINWYDFQPHLLPNVEFVGLACRPAXELPKEMEFPVQS 299
 CC Db 241 LGRPTTLETWGAEMWLRITVWDEBPFRPLPVNDVFGALHCRPAKPLPEMEBEPVQS 300
 CC QY 300 GENGVVFSLSSMNTSEEPANVYASALAKIPQKVLRFPGNKPDTGLNTRLYKATPQ 359
 CC Db 301 GENGIIVFSLGSMISNNEBSANNIASALAOIPQVLNRFPGKPKNTLGSNTRLYKATPQ 360
 CC QY 360 NDILGHPTKAFITHGNGIYEALYHGVPVWGVIPFGDQDNIAHMKAKGAIVEINPKT 419

Db 361 NDILGHPTKAFITHGNGIYEALYHGVPVWGVIPFGDQDNIAHMKAKGAALSDVIRT 420
 CC QY 420 MTSDDLRLARTVYITDSYKENMRSLRIHDDPVKRLDAVFIETVMHKKAKHLRSA 479
 CC Db 421 MSSDDLNLAKSYINDIYENKMLSRHHDDPVKRLDAVFIETVMHKKAKHLRVA 480
 CC QY 480 AHDLPWFQHSIDVIGFLCVATAIPLFTKCFPLFSQCKFNKRRIKRE 529
 CC Db 481 AHDLPWFQHSIDVIGFLCVATAIPLFTKCFPLFSQCKFNKRRIKRE 530
 CC
 CC RESULT 9
 CC IDBFF HUMAN STANDARD; PRT; 530 AA.
 CC AC P54855; P23765; Q9URK63;
 CC DT 01-NOV-1991 (Rel. 20, Created)
 CC DT 01-OCT-1996 (Rel. 34, Last sequence update)
 CC DT 15-SEP-2003 (Rel. 42, Last annotation update)
 CC DE UDP-glucuronosyltransferase 2B15 precursor, microsomal (EC 2.4.1.17)
 CC (UDPGT) (UDPGT-3) (HUG4).
 CC GN UGT2B15 OR UGT2B8.
 CC OS Homo sapiens (Human).
 CC OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 CC OX NCBI_TaxId=9606;
 CC RN [1]
 CC RP SEQUENCE FROM N.A.
 CC RC TISSUE=Liver;
 CC RX MEDLINE=95136867; PubMed=7835232;
 CC RA Green M.P., Oltu E.M., Tepfly T.R.;
 CC RT "Stable expression of a human liver UDP-glucuronosyltransferase
 CC (UGT2B15) with activity toward steroid and xenobiotic substrates.";
 CC RL Drug Metab. Dispos. 22:799-805(1994).
 CC RN [2]
 CC RP SEQUENCE OF 8-530 FROM N.A., AND PARTIAL SEQUENCE.
 CC RC TISSUE=Liver;
 CC RX MEDLINE=90343358; PubMed=2116769;
 CC RA Coffman B.L., Tepfly T.R., Irsheid Y.M., Green M.D., Smith C.,
 CC Jackson M.R., Wooster R., Burchell B.;
 CC RT "Characterization and primary sequence of a human hepatic microsomal
 CC estriol UDP-glucuronosyltransferase.";
 CC RL Arch. Biochem. Biophys. 281:170-175(1990).
 CC RN [3]
 CC RP SEQUENCE FROM N.A. AND VARIANT TYR-85.
 CC RX MEDLINE=97439504; PubMed=9295060;
 CC RA Levesque E., Beaulieu M., Green M.D., Tepfly T.R., Belanger A.,
 CC Hum D.W.;
 CC RT "Isolation and characterization of UGT2B15(Y85): a
 CC UDP-glucuronosyltransferase encoded by a polymorphic gene.";
 CC RL Pharmacogenetics 7:317-325(1997).
 CC RN [4]
 CC RP SEQUENCE FROM N.A.
 CC RC TISSUE=Liver;
 CC RX Owens I.S.;
 CC RA Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
 CC CC -1- FUNCTION: UDPGTs are of major importance in the conjugation and
 CC subsequent elimination of potentially toxic xenobiotics and
 CC endogenous compounds. This isozyme displays activity toward
 CC several classes of xenobiotic substrates, including simple
 CC phenolic compounds, 7-hydroxylated coumarins, flavonoids,
 CC anthraquinones, and certain drugs and their hydroxylated
 CC metabolites. It also catalyzes the glucuronidation of endogenous
 CC estrogens and androgens.
 CC CC -1- CATALYTIC ACTIVITY: UDP-glucuronate + acceptor = UDP + acceptor
 CC beta-D-glucuronoside.
 CC CC -1- SUBCELLULAR LOCATION: Microsomal.
 CC CC -1- TISSUE SPECIFICITY: Expressed in many tissues.
 CC CC -1- SIMILARITY: BELONGS TO THE UDP-GLYCOSYLTRANSFERASE FAMILY.
 CC CC -1- CAUTION: REF. 2 THOUGH THAT THIS WAS A SEPARATE FORM (UGT2B8). THE
 CC NAME UGT2B8 NOW BEEN REUSED FOR A RAT ENZYME.
 CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.emb-emb.ch/announce/>
 CC or send an email to license@emb-emb.ch).

DR EMBL: U08854; AAC50077.1; -;
 DR EMBL: U06641; AAA83406.1; -;
 DR EMBL: AF180322; AAD55093.1; -;
 DR EMBL: AF548389; AAA40695.1; -;
 DR PIR: A48633; A48633.
 DR Genew; HGNC:12546; UGT2B15.
 DR MIM: 600069; -;
 DR GO: GO:0008202; P:steroid metabolism; TAS.
 DR GO: GO:0006805; P:xenobiotic metabolism; TAS.
 DR InterPro; IPR002213; UDP_gluco_trans.
 DR Pfam; PF00201; UDPGT; 1.
 DR PROSITE; PS00375; UDPGT; 1.
 KW Transferase; Glycosyltransferase; Glycoprotein; Transmembrane; Signal;
 KW Multigene family; Microsome; Polymorphism.
 FT SIGNAL 1 23
 FT CHAIN 1 23
 FT TRANSMEM 24 530
 FT CARBOHYD 495 515
 FT CARBOHYD 65 515
 FT CARBOHYD 316 515
 FT CARBOHYD 483 515
 FT VARIANT 85 85
 FT FT
 FT CONFLICT 119 119
 FT CONFLICT 145 145
 FT CONFLICT 150 155
 FT CONFLICT 162 162
 FT CONFLICT 165 165
 FT CONFLICT 170 177
 FT CONFLICT 181 203
 FT CONFLICT 203 203
 FT CONFLICT 293 293
 FT CONFLICT 401 401
 FT CONFLICT 443 443
 FT CONFLICT 501 501
 FT CONFLICT 523 523
 SQ SEQUENCE 530 AA; 60961 MW; 23DDA4B9687CB4 CRC64;
 Query Match 77.1%; Score 2158.5; DB 1; Length 530;
 Best Local Similarity 74.5%; Pred. No. 7.2e-156;
 Matches 395; Conservative 60; Mismatches 74; Indels 1; Gaps 1;
 Oy 1 MSMKWTALLILQLSCYFSSGCGKVLWPTFFSHMNIKTLLDELVORGHEVTVLASSA 60
 Db 1 MSKMTSYPLILQLSCYFSSGCGKVLWPTFFSHMNIKTLLDELVORGHEVTVLASSA 60
 Oy 61 SISDPNSPSTLKFEVYVSLTKTEFEDIIKQLVKRA-ELPKDTFMSYFSGVGEIMWTF 119
 Db 61 STLVASKSSAIKLEVYPTSLTKNLEDSLKIILRLMVIYSGSKTFMWSYFSGVGEIMWTF 120
 Oy 120 NDILAFKCDVYNSKKMKKQOESFEDVILADVPPEPPELAEILKIPVVSILRSPGYA 179
 Db 121 YDYSNKLCKDAVLNKKMKKQOESFEDVILADVPPEPPELAEILKIPVVSILRSPGYA 180
 Oy 180 IEKHSGLLFPSPYVYVMSSELSDMTFIERVKNNIYVLFEPFQIFDMKMDQFYSYEV 239
 Db 181 FEKNGGGLFPSPYVYVMSSELSDMTFIERVKNNIYVLFEPFQIFDMKMDQFYSYEV 240
 Oy 240 LGRPTTLEETAKADWILIRNYWDFQPHPLLPNVEFVGGLCKCPAKPLPKEMEEFYOSS 299
 Db 241 LGRPTTLEETAKADWILIRNYWDFQPHPLLPNVEFVGGLCKCPAKPLPKEMEEFYOSS 300
 Oy 300 GENGIVVSLGSMVNTSEERANVASALAKIPKVLWRFDQNKDPTGLNRLTKYKIPQ 359
 Db 301 GENGIVVSLGSMVNTSEERANVASALAKIPKVLWRFDQNKDPTGLNRLTKYKIPQ 360
 Oy 360 NDLLGHPTKAPITLHGNNGIYEAIVHGVPMVGVPIFGQDLNIAHMKAKGAIVEINFKT 419

Db 361 NDLLGHPTKAPITLHGNNGIYEAIVHGVPMVGVPIFGQDLNIAHMKAKGAALSVIRT 420
 Oy 420 MTSFDLRLARTVTTDSYSENNMRSLRIHHDDQVRLDRAVFIETFMHKGAKHLRSA 479
 Db 421 MSSDILNALKSVINDVYENVKLSRIHHDDQVRLDRAVFIETFMHKGAKHLRSA 480
 Oy 480 AHDLTWFQHSYIDVIGFLNCVATAIFLFTKCFPLSCQKFNKTKRIEKRE 529
 Db 481 AHNLTWIOYHSLDVIAFLACVATVIFITKFCIFCRKLAKTKKKKKD 530

RESULT 10
 UDBK_MACFA
 ID UDBK_MACFA STANDARD; PRT; 530 AA.
 AC 077649;
 DT 15-DEC-1998 (Rel. 37, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE UDP-glycuronosyltransferase 2B20 precursor, microsomal (EC 2.4.1.17)
 DE (UNPRT).
 GN UGT2B20.
 OS Macaca fascicularis (Crab eating macaque) (Cynomolgus monkey).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
 OC Cercopithecoidea; Macaca.
 CX NCBI_TaxID=9541;
 RN [1]
 RP SEQUENCE FROM N.A. AND CHARACTERIZATION.
 RC TISSUE=Liver, and Prostate.
 RX MEDLINE=99112924; PubMed=8995103;
 RA Barbier O., Belanger A., Hum D.W.;
 RT "Cloning and characterization of a simian UDP-glycuronosyltransferase
 enzyme UGT2B20, a novel C19 steroid-conjugating protein.";
 RL Biochem. J. 337:567-574(1999).
 CC -1- FUNCTION: UDPGs are of major importance in the conjugation and
 CC subsequent elimination of potentially toxic xenobiotics and
 CC endogenous compounds. This isozyme has glucuronidating capacity
 CC with androgens, such as testosterone, dihydrotestosterone (DHT)
 CC and alpha-diol. It is also active on catecholoeestrogens including
 CC 1,3,5,10-oestradiene-3,4-diol-17-one.
 CC -1- CATALYTIC ACTIVITY: UDP-glucuronate + acceptor = UDP + acceptor
 CC beta-D-glucuronoside.
 CC -1- SUBCELLULAR LOCATION: Microsomal.
 CC -1- SIMILARITY: BELONGS TO THE UDP-GLYCOSYLTRANSFERASE FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.emb-emb.ch/announce/>
 CC or send an email to license@emb-emb.ch).

Query Match 75.1%; Score 2100.5; DB 1; Length 530;
 Best Local Similarity 73.1%; Pred. No. 1.8e-151;
 Matches 386; Conservative 60; Mismatches 81; Indels 1; Gaps 1;

```

QY 1 MSKMTSALLIOLQSCYFSSGCGKVLWPTFESHMMNITLDELVQGHVTVLASSA 60
D 1 MSKMTSVFLLIOLQSCYFSSGCGKVLWPTFESHMMNITLDELVQGHVTVLASSA 60
QY 61 SISDPNSPTLKEEYVPVSLTKTEFEDIIKQLYKMA-ELPKDTFMSYSQVOEIMTF 119
D 61 STFNVDKSSSAIKKEEYVPTSLTKMDMEDSLMKLIDITWYSISNSTFLSYFSKLOELCMEY 120
QY 120 NDILRKCCKIIVSKMKLKOESRFPVLADAVFPFEGELLAEELKI PFYYSRFSFGYA 179
D 121 YYISEKCKOAVLNKKMTKTKETKFDVILLADALNPGELAELEFNI PFYYSLEFVGYT 180
QY 180 IEKSGGLPEPSPVYVPMSELSDQMTFIERVKMNIYLYFEFQOIFDMKMDQFYSEV 239
D 181 FEKNGGFLPFPSPVPMSELSDQMTFIERVKMNIYLYFEFQOIFDMKMDQFYSEV 240
QY 240 LGRPTLSEFMKADIVLIRNYWDPORPHPLPNVEVVGGLHCKPAKPLPKMEEFVQSS 299
D 241 LGRPTLSEFMKADIVLIRNYWDPORPHPLPNVEVVGGLHCKPAKPLPKMEEFVQSS 300
QY 300 GENGWVFSLSGWSVNTSEERANVIASALAKIPQKVLMPDGNKPDGLINTRLYKMIPO 359
D 301 GENGWVFSLSGWSVNTSEERANVIASALAKIPQKVLMPDGNKPDGLINTRLYKMIPO 360
QY 360 NDILGHPKTAFTTHGANGIYEAIYHGVPMVGPVIFGDLNIAHMKAKAAVEINFKT 419
D 361 NDILGHPKTAFTTHGANGIYEAIYHGVPMVGPVIFGDLNIAHMKAKAAVEINFKT 420
QY 420 MTSDDLRLARTVITDSSYKENAMRLSRHHDDQVPLDRAVWIEFVMRHKGAHLRSA 479
D 421 MSSDDLRLARTVITDSSYKENAMRLSRHHDDQVPLDRAVWIEFVMRHKGAHLRSA 480
QY 480 AHDLTWFQHSIDVIGFLTCVATAIFLFTKCFLEFSQCKNKRKTRE 527
D 481 AHDLTWFQHSIDVIGFLTCVATAIFLFTKCFLEFSQCKNKRKTRE 528

RESULT 11
UDBE_RABIT STANDARD; PRT; 530 AA.
ID UDBE_RABIT STANDARD; PRT; 530 AA.
AC P36513;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE UDP-glucuronosyltransferase 2B14 precursor, microsomal (EC 2.4.1.17)
DE (UDPGT) (EGT12).
GN UGT2B14.
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX NCBI_TaxID=9986;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=New Zealand white; TISSUE=Liver;
RX MEDLINE=93315511; PubMed=8325897;
RA Turkey R.H., Pendurthi U.R., Nguyen N.T., Green M.D., Tephly T.R.;
RT "Cloning and characterization of rabbit liver UDP-
RT glucuronosyltransferase cDNAs. Developmental and inducible expression
RT of 4-hydroxydiphenyl UGT2B13."
RL J. Biol. Chem. 268:15260-15266(1993).
CC -1- FUNCTION: UDPGT IS OF MAJOR IMPORTANCE IN THE CONJUGATION AND
CC SUBSEQUENT ELIMINATION OF POTENTIALLY TOXIC XENOBIOTICS AND
CC ENDOGENOUS COMPOUNDS.
CC -1- CATALYTIC ACTIVITY: UDP-glucuronate + acceptor = UDP + acceptor
CC beta-D-glucuronoside.
CC -1- SUBCELLULAR LOCATION: Microsomal.
CC -1- DEVELOPMENTAL STAGE: EXPRESSED PRIMARILY IN ADULT RABBITS.
CC -1- SIMILARITY: BELONGS TO THE UDP-GLUCOSYLTRANSFERASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way

```

```

CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.ebi.ac.uk/announcements/
CC or send an email to license@ebi.ac.uk).
CC -----
DR EMBL; L01082; AAA18021.1; -.
DR PIR; C47113; C47113.
DR InterPro; IPR002213; UDP_gluco_trans.
DR Pfam; PF00201; UDPGT; 1.
DR PROSITE; PS00375; UDPGT; 1.
KW Transferrase; Glycosyltransferase; Glycoprotein; Transmembrane; Signal;
KW Multigene family; Microsome.
FT SIGNAL 1 24
FT CHAIN 25 530 UDP-GLUCURONOSYLTRANSFERASE 2B14.
FT TRANSMEM 494 510 POTENTIAL.
FT CARBOHYD 134 134 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 316 316 N-LINKED (GLCNAC...) (POTENTIAL).
SQ SEQUENCE 530 AA; 60664 MW; CEAA4BF25B53CB5 CRC64;

Query Match 71.2%; Score 1991.5; DB 1; Length 530;
Best Local Similarity 67.9%; Pred. No. 3,2e-143;
Matches 360; Conservative 79; Mismatches 90; Indels 1; Gaps 1;

QY 1 MSKMTSALLIOLQSCYFSSGCGKVLWPTFESHMMNITLDELVQGHVTVLASS 59
D 1 MSKMTSVFLLIOLQSCYFSSGCGKVLWPTFESHMMNITLDELVQGHVTVLASS 60
QY 60 ASISFPNSSTLKEEYVPVSLTKTEFEDIIKQLYKMA-ELPKDTFMSYSQVOEIMTF 119
D 61 ASISFPNSSTLKEEYVPVSLTKTEFEDIIKQLYKMA-ELPKDTFMSYSQVOEIMTF 120
QY 120 NDILRKCCKIIVSKMKLKOESRFPVLADAVFPFEGELLAEELKI PFYYSRFSFGYA 179
D 121 NDILRKCCKIIVSKMKLKOESRFPVLADAVFPFEGELLAEELKI PFYYSRFSFGYT 180
QY 180 IEKSGGLPEPSPVYVPMSELSDQMTFIERVKMNIYLYFEFQOIFDMKMDQFYSEV 239
D 181 IEKSGGLPEPSPVYVPMSELSDQMTFIERVKMNIYLYFEFQOIFDMKMDQFYSEV 240
QY 240 LGRPTLSEFMKADIVLIRNYWDPORPHPLPNVEVVGGLHCKPAKPLPKMEEFVQSS 299
D 241 LGRPTLSEFMKADIVLIRNYWDPORPHPLPNVEVVGGLHCKPAKPLPKMEEFVQSS 300
QY 300 GENGWVFSLSGWSVNTSEERANVIASALAKIPQKVLMPDGNKPDGLINTRLYKMIPO 359
D 301 GENGWVFSLSGWSVNTSEERANVIASALAKIPQKVLMPDGNKPDGLINTRLYKMIPO 360
QY 360 NDILGHPKTAFTTHGANGIYEAIYHGVPMVGPVIFGDLNIAHMKAKAAVEINFKT 419
D 361 NDILGHPKTAFTTHGANGIYEAIYHGVPMVGPVIFGDLNIAHMKAKAAVEINFKT 420
QY 420 MTSDDLRLARTVITDSSYKENAMRLSRHHDDQVPLDRAVWIEFVMRHKGAHLRSA 479
D 421 MTSDDLRLARTVITDSSYKENAMRLSRHHDDQVPLDRAVWIEFVMRHKGAHLRSA 480
QY 480 AHDLTWFQHSIDVIGFLTCVATAIFLFTKCFLEFSQCKNKRKTRE 529
D 481 AHDLTWFQHSIDVIGFLTCVATAIFLFTKCFLEFSQCKNKRKTRE 530

RESULT 12
UDBE_RAT STANDARD; PRT; 529 AA.
ID UDBE_RAT STANDARD; PRT; 529 AA.
AC P09875;
DT 01-MAR-1989 (Rel. 10, Created)
DT 01-MAR-1989 (Rel. 10, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE UDP-glucuronosyltransferase 2B1 precursor, microsomal (EC 2.4.1.17)
DE (UDPGT) (UDPGTR-2).
GN UGT2B1.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;

```

```

RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RX MEDLINE=66196018; PubMed=3084479;
RA McKenzie P.I.;
RT "Rat liver UDP-glucuronosyltransferase. Sequence and expression of a
  cDNA encoding a phenobarbital-inducible form.";
RL J. Biol. Chem. 261:619-6125(1986).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=90293083; PubMed=2113533;
RA McKenzie P.I.; Roddoun L.;
RT "Organization of the rat UDP-glucuronosyltransferase, UDPGtr-2, gene.
  and characterization of its promoter.";
RL J. Biol. Chem. 265:11328-11333(1990).
CC -1- FUNCTION: UDPGT IS OF MAJOR IMPORTANCE IN THE CONJUGATION AND
  SUBSEQUENT ELIMINATION OF POTENTIALLY TOXIC XENOBIOTICS AND
  ENDOGENOUS COMPOUNDS.
CC -1- CATALYTIC ACTIVITY: UDP-glucuronate + acceptor = UDP + acceptor
  beta-D-glucuronoside.
CC -1- SUBCELLULAR LOCATION: Microsomal.
CC -1- INDUCTION: By phenobarbital.
CC -1- SIMILARITY: BELONGS TO THE UDP-GLYCOSYLTRANSFERASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
  between the Swiss Institute of Bioinformatics and the EMBL outstation -
  the European Bioinformatics Institute. There are no restrictions on its
  use by non-profit institutions as long as its content is in no way
  modified and this statement is not removed. Usage by and for commercial
  entities requires a license agreement (See http://www.isb-sib.ch/announce/
  or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M13506; AAA42313.1; -
DR EMBL; M35086; AAA42310.1; JOINED.
DR EMBL; M35202; AAA42310.1; JOINED.
DR EMBL; M35080; AAA42310.1; JOINED.
DR EMBL; M35082; AAA42310.1; JOINED.
DR EMBL; M35083; AAA42310.1; JOINED.
DR PIR; A42233; A42233.
DR InterPro; IPR002213; UDP_gluco_trans.
DR Pfam; PF00201; UDPGT; 1.
DR PROSITE; PS00375; UDPGT; 1.
KW Transferase; Glycosyltransferase; Glycoprotein; Transmembrane; Signal;
  Multigene family; Microsome.
FT SIGNAL 1 23 BY SIMILARITY.
FT CHAIN 24 529 UDP-GLUCURONOSYLTRANSFERASE 2B1.
FT TRANSMEM 494 510 POTENTIAL.
FT CARBOHYD 134 134 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 316 316 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 529 AA; 60484 MW; 14DF0224BF1C3044 CRC64;

Query Match 70.5%; Score 1973.5; DB 1; Length 529;
Best Local Similarity 68.4%; Pred. No. 7.3e-142;
Matches 361; Conservative 68; Mismatches 98; Indels 1; Gaps 1;

```

```

Db 241 LGRPTLTTEMKKADILWIRTFWDLFFPHPLPNFDFVGLHCKPAKPLPREKEEFVQSS 300
Oy 300 GENGCVVFSLSGWSVNTSEERANVIASALAKIPQKVLMPREDGNKPTDGLNTRLYKMIPO 359
Db 301 GENGCVVFSLSGWSVNTSEERANVIASALAKIPQKVLMPREDGNKPTDGLNTRLYKMIPO 360
Oy 360 NDILGHPKTAFTTHGGMNCIYEAIYHGVPMVGVPIFGDLDNIAMKAGAVEINFKT 419
Db 361 NDILGHPKTAFTTHGGMNCIYEAIYHGVPMVGVPIFGDLDNIAMKAGAVEINFKT 420
Oy 420 MTSDDLRAIRTYITDSYKENMRSLRIHDDPVKRLDAVAFIEFVMRHKGAKHLSA 479
Db 421 LSTGLTALAKIWNDSYKENMRSLRIHDDPVKRLDAVAFIEFVMRHKGAKHLSA 480
Oy 480 AHDLTWFQHSIDVIGFLTCVATAPLFTKCFPLSCCKENKTRIKER 527
Db 481 LHDLSWFQHSIDVIGFLTCVATAPLFTKCFPLSCCKENKTRIKER 528

```

RESULT 13

```

ID _UDBG_RABIT STANDARD; PRT; 523 AA.
AC 019103;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE UDP-glucuronosyltransferase 2B16 precursor, microsomal (EC 2.4.1.17)
  (UDPGtr) (Fragment).
GN UGT2B16.
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Leporidae; Oryctolagus.
OX NCBI_Taxid=9986;

```

```

RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97166166; PubMed=9013565;
RA Li Q., Lou X., Peyronneau M.-A., Straub P.O., Tukey R.H.;
RT "Expression and functional domains of rabbit liver UDP-
  glucuronosyltransferase 2B16 and 2B13.";
RL J. Biol. Chem. 272:3272-3279(1997).
CC -1- FUNCTION: UDPGT IS OF MAJOR IMPORTANCE IN THE CONJUGATION AND
  SUBSEQUENT ELIMINATION OF POTENTIALLY TOXIC XENOBIOTICS AND
  ENDOGENOUS COMPOUNDS. ACTS ON SMALL PHENOLIC AGENTS SUCH AS 2-
  NAPHTHOL AND 4-METHYLBELIFERONE AS WELL AS BULKY PHENOLIC
  COMPOUNDS LIKE 2-HYDROXY- AND 4-HYDROXYBIPHENYL. IN CONTRAST TO
  2B13 IT IS ACTIVE TOWARD 4-HYDROXYESTERONE.
CC -1- CATALYTIC ACTIVITY: UDP-glucuronate + acceptor = UDP + acceptor
  beta-D-glucuronoside.
CC -1- SUBCELLULAR LOCATION: Microsomal.
CC -1- DEVELOPMENTAL STAGE: EXPRESSED PRIMARILY IN ADULT RABBITS.
CC -1- SIMILARITY: BELONGS TO THE UDP-GLYCOSYLTRANSFERASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
  between the Swiss Institute of Bioinformatics and the EMBL outstation -
  the European Bioinformatics Institute. There are no restrictions on its
  use by non-profit institutions as long as its content is in no way
  modified and this statement is not removed. Usage by and for commercial
  entities requires a license agreement (See http://www.isb-sib.ch/announce/
  or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U72742; AAB71494.1; -
DR InterPro; IPR002213; UDP_gluco_trans.
DR Pfam; PF00201; UDPGT; 1.
DR PROSITE; PS00375; UDPGT; 1.
KW Transferase; Glycosyltransferase; Glycoprotein; Transmembrane; Signal;
  Multigene family; Microsome.
FT SIGNAL 1 16
FT CHAIN 17 523 UDP-GLUCURONOSYLTRANSFERASE 2B16.
FT TRANSMEM 487 503 POTENTIAL.
FT CARBOHYD 309 309 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 523 AA; 60077 MW; 22408BE768ED3DE CRC64;

```


DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE UDP-glucuronosyltransferase 2B12 precursor, microsomal (EC 2.4.1.17)
DE (UDPGLT).
GN UGT2B12.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.
OX NCBI_TaxId=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Sprague-Dawley; TISSUE=Kidney, and Liver;
RX MEDLINE=96032669; PubMed=7574722;
RA Green M.D., Clarke D.J., Ocaru E.M., Styczynski P.B.,
RA Jackson M.R., Burchell B., Teply T.R.;
RT "Cloning and expression of a rat liver phenobarbital-inducible UDP-
RT glucuronosyltransferase (2B12) with specificity for monoterpenoid
RT alcohols";
RL Arch. Biochem. Biophys. 322:460-468(1995).
RN [2]
RP SEQUENCE OF 24-38, AND CHARACTERIZATION.
RC STRAIN=Mistar; TISSUE=Liver;
RX MEDLINE=91312327; PubMed=1906977;
RA Styczynski P.B., Green M.S., Puig J., Coffman B.L., Teply T.R.;
RT "Purification and properties of a rat liver phenobarbital-inducible
RT 4-hydroxyphenyl UDP-glucuronosyltransferase";
RL Mol. Pharmacol. 40:80-84(1991).
CC -1- FUNCTION: UDPGT IS OF MAJOR IMPORTANCE IN THE CONJUGATION AND
CC SUBSEQUENT ELIMINATION OF POTENTIALLY TOXIC XENOBIOTICS AND
CC ENOGENOUS COMPOUNDS. CATALYZES THE TRANSFER OF GLUCURONIC ACID
CC FROM UDP-GLUCURONIC ACID TO VARIOUS AGLYCONE MOLECULES. CATALYZES
CC THE GLUCONIDATION OF MONOTERPENOID ALCOHOLS, SUCH AS (-)-
CC BONEOL, (+)-MENTHOL, AND (-)-NOFOL. IN ADDITION, A NUMBER OF
CC SIMPLE PHENOLIC COMPOUNDS, SUCH AS HYDROXYBIPHENYLS, 7-
CC HYDROXYLATED COUMARINS, P-NITROPHENOL, AND FOOD-DERIVED SUBSTANCES
CC (E.G., NAINEBIN AND EUGENOL), AND 4-METHYLBELLIFERONE ARE ALSO
CC SUBSTRATES.
CC -1- CATALYTIC ACTIVITY: UDP-glucuronate + acceptor = UDP + acceptor
CC beta-D-glucuronoside.
CC -1- SUBCELLULAR LOCATION: Microsomal.
CC -1- TISSUE SPECIFICITY: LIVER. LOWER LEVELS SEEN IN THE KIDNEY AND
CC TESTIS.
CC -1- INDUCTION: By phenobarbital.
CC -1- PTM: N-GLYCOSYLATED (PROBABLY).
CC -1- POLYMORPHISM: THE SEQUENCE SHOWN IS THAT OF THE LIVER ISOZYME.
CC THE KIDNEY ISOFORMS DIFFERS IN 12 POSITIONS.
CC -1- SIMILARITY: BELONGS TO THE UDP-GLYCOSYLTRANSFERASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib.ch).
CC -----
DR EMBL: U06273; AAA83404.1; -
DR EMBL: U06274; AAA83405.1; -
DR PIR: S68200; S68200.
DR InterPro: IPR002213; UDP_gluco_trans.
DR Pfam: PF00201; UDPGT_1.
DR ProSite: PS00375; UDPGT_1.
KW Transferase; Glycosyltransferase; Glycoprotein; Transmembrane; Signal;
KW Multigene family; Microsome.
FT SIGNAL 1 23
FT CHAIN 24 530 UDP-GLUCURONOSYLTRANSFERASE 2B12.
FT TRANSMEM POTENTIAL.
FT CARBOHYD 316 316 N-LINKED (GLCNAC. .) (POTENTIAL).
FT VARIANT 2 2 S -> P (IN KIDNEY).
FT VARIANT 61 61 F -> S (IN KIDNEY).
FT VARIANT 71 71 D -> H (IN KIDNEY).
FT VARIANT 95 95 N -> S (IN KIDNEY).
FT VARIANT 183 183 Q -> K (IN KIDNEY).
FT VARIANT 346 346 P -> T (IN KIDNEY).

FT VARIANT 398 398 A -> G (IN KIDNEY).
FT VARIANT 414 415 VE -> AT (IN KIDNEY).
FT VARIANT 433 433 V -> D (IN KIDNEY).
FT VARIANT 475 475 K -> L (IN KIDNEY).
FT VARIANT 488 488 Q -> L (IN KIDNEY).
SQ SEQUENCE 530 AA; 61060 MW; D49313CE3E6D5BED CRC64;
Query Match 69.0%; Score 1930.5; DB 1; Length 530;
Best Local Similarity 66.8%; Pred. No. 1.3e-138;
Matches 354; Conservative 73; Mismatches 102; Indels 1; Gaps 1;
QY 1 MSMTATALLLIQLSCYFSSGCGKVLVMPFESHMNKTILDELVQRGEVTVLASSA 60
DB 1 MSGKIALALLLIQISFCFKSGCKGLVMPWEYSHMNKILIELVQKGEVTVLRPSA 60
QY 61 SISDPSPSPITKEVYVSLTKEFDIIKQVKKRA-ELPKDTPMSYQVOEIMWT 119
DB 61 FVLDPKETSCLKVTPTPTSSSHDLNPTFRFVNVVTELPRTCLSYFLYLDOTIDY 120
QY 120 NDILRKCKDIVSNKMLKLOESRPDVVLADAVFPGELALDLKIPVYSLFRSPGYA 179
DB 121 SDYCLTYCKEAVSNKQMTKLOESKFDVFSDAIPGGLALDLKIPVLSLFRSPGYT 180
QY 180 IEKHSGLLEPPSYVPMVMSLSQMTPIERVKMIVLYIFEFWQIFDKKKDQFISEV 239
DB 181 IEQYIGVLPSPSYVPMIFSGLAQMTFIERVHNMICLYEDFWQIFREKKMDPFYSKT 240
QY 240 LGRPTTSETMAKADILIRVYDFOPPHLLPVVEFVGJHCAPKPLKEMEFPVQSS 299
DB 241 LGRPTTALETGKAEMLIRYVLEPPIPSVVDYIGJHCAPKPLKPDIEDFVQSS 300
QY 300 GENGVVFSLSGMSVNTSEERANVYASALAKIPQKVLRFPGNKPDTLGLNTRYKNIPO 359
DB 301 GEHGVVFSLSGMSVNTSEERANVYASALAKIPQKVLRFPGNKPDTLGLNTRYKNIPO 360
QY 360 NDILGHPKTKAFITHGNGNGIYEAIVGVPMVGVPIFGDQLDNIAHMKAKAAVEINFT 419
DB 361 NDILGHPKTKAFVTHGNGNGIYEAIVGHGIPMIGIPLFAEQHDNIAHMAKAAVEINFT 420
QY 420 MTSDDLRLALRTVTTDSYKENAMRLSRHHDOVKPLDRAVVFIEFVRRKGAHRLSA 479
DB 421 MSKSDLLNAAEEVVDNPPYKNAAMLSTIHHDOPTKPLDRAVVFIEFVRRKGAHRLSL 480
QY 480 AHDLTWFOHYSIDVIGFLTCVATAIFLTKCFLFSCQKFNKTRIEKRE 529
DB 481 GNLFPWQYHSLDVIGFLSCVAATVIALKCFIVVRFVKEKKTKNE 530

Search completed: December 5, 2003, 09:48:50
Job time : 19 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 5, 2003, 09:47:25 / Search time 21 Seconds
(without alignments)
1065.831 Million cell updates/sec

Title: US-09-980-729B-5
Perfect score: 2798
Sequence: 1 MSKMTSALLLIOLSCYFSS.....KCFLLFSCQKFKTRKRE 529

Scoring table: BIOSUM62
Gapop 10.0, Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA:
1: /cgn2_6/ptodata/1/1aa/5A_COMB.pep.*
2: /cgn2_6/ptodata/1/1aa/5B_COMB.pep.*
3: /cgn2_6/ptodata/1/1aa/6A_COMB.pep.*
4: /cgn2_6/ptodata/1/1aa/6B_COMB.pep.*
5: /cgn2_6/ptodata/1/1aa/PTUS_COMB.pep.*
6: /cgn2_6/ptodata/1/1aa/backfile1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|--------|-------------|--------|----|-------------------|
| 1 | 2602.5 | 93.0 | 528 | 4 | US-09-356-806-8 |
| 2 | 2322 | 83.0 | 524 | 4 | US-09-356-806-40 |
| 3 | 2169.5 | 77.5 | 530 | 3 | US-09-180-852-2 |
| 4 | 2158.5 | 77.1 | 530 | 4 | US-09-356-806-113 |
| 5 | 1874.5 | 67.0 | 454 | 4 | US-09-813-918-2 |
| 6 | 1243 | 44.4 | 288 | 4 | US-09-813-918-3 |
| 7 | 1196.5 | 42.8 | 531 | 5 | PCT-US92-00282-6 |
| 8 | 1130.5 | 40.4 | 533 | 5 | PCT-US92-00282-3 |
| 9 | 1126 | 40.2 | 534 | 5 | PCT-US92-00282-4 |
| 10 | 1064 | 38.0 | 531 | 5 | PCT-US92-00282-7 |
| 11 | 1045.5 | 37.4 | 529 | 5 | PCT-US92-00282-7 |
| 12 | 833 | 29.8 | 197 | 4 | US-09-813-918-4 |
| 13 | 765.5 | 27.4 | 245 | 4 | US-09-305-856B-18 |
| 14 | 403 | 14.4 | 515 | 3 | US-08-942-012B-32 |
| 15 | 393 | 14.0 | 129 | 4 | US-09-370-838-36 |
| 16 | 379 | 13.5 | 488 | 3 | US-08-942-012B-29 |
| 17 | 379 | 13.5 | 488 | 3 | US-08-942-012B-30 |
| 18 | 374 | 13.4 | 460 | 3 | US-08-942-012B-33 |
| 19 | 372 | 13.3 | 288 | 4 | US-09-305-856B-2 |
| 20 | 372 | 13.3 | 288 | 4 | PCT-US92-00282-19 |
| 21 | 362.5 | 13.0 | 289 | 4 | US-09-305-856B-8 |
| 22 | 360.5 | 12.9 | 289 | 4 | US-09-305-856B-6 |
| 23 | 360.5 | 12.9 | 289 | 5 | PCT-US92-00282-13 |
| 24 | 358.5 | 12.8 | 289 | 4 | US-09-305-856B-4 |
| 25 | 358.5 | 12.8 | 289 | 4 | PCT-US92-00282-15 |
| 26 | 354.5 | 12.7 | 289 | 5 | PCT-US92-00282-11 |
| 27 | 346 | 12.4 | 310 | 4 | US-09-305-856B-14 |

| | | | | | | |
|----|-------|------|-----|---|-------------------|--------------------|
| 28 | 338 | 12.1 | 515 | 3 | US-08-942-012B-24 | Sequence 24, Appl |
| 29 | 321.5 | 11.5 | 317 | 4 | US-09-305-856B-12 | Sequence 12, Appl |
| 30 | 316 | 11.3 | 287 | 4 | US-09-305-856B-10 | Sequence 10, Appl |
| 31 | 306.5 | 11.0 | 286 | 5 | PCT-US92-00282-9 | Sequence 9, Appl |
| 32 | 297.5 | 10.6 | 98 | 5 | PCT-US92-00282-26 | Sequence 26, Appl |
| 33 | 296.5 | 10.6 | 253 | 4 | US-09-305-856B-16 | Sequence 16, Appl |
| 34 | 294.5 | 10.5 | 506 | 3 | US-08-942-012B-26 | Sequence 26, Appl |
| 35 | 287 | 10.3 | 506 | 3 | US-08-942-012B-25 | Sequence 25, Appl |
| 36 | 287 | 10.3 | 506 | 6 | 5180581-2 | Patent No. 5180581 |
| 37 | 271 | 9.7 | 58 | 2 | US-08-466-583-9 | Sequence 9, Appl |
| 38 | 271 | 9.7 | 58 | 4 | US-08-265-427-9 | Sequence 9, Appl |
| 39 | 271 | 9.7 | 58 | 5 | PCT-US95-07820-9 | Sequence 9, Appl |
| 40 | 261 | 9.3 | 493 | 3 | US-08-942-012B-28 | Sequence 28, Appl |
| 41 | 257 | 9.2 | 52 | 2 | US-08-466-583-7 | Sequence 7, Appl |
| 42 | 257 | 9.2 | 52 | 4 | US-08-265-427-7 | Sequence 7, Appl |
| 43 | 257 | 9.2 | 52 | 5 | PCT-US95-07820-7 | Sequence 7, Appl |
| 44 | 250.5 | 9.0 | 489 | 3 | US-08-942-012B-31 | Sequence 31, Appl |
| 45 | 241.5 | 8.6 | 491 | 3 | US-08-942-012B-27 | Sequence 27, Appl |

ALIGNMENTS

| | | | |
|--|------------------|--|-------------------------------------|
| RESULT 1 | | US-09-356-806-8 | Sequence-B: Application US/09356806 |
| | | Patent No. 6586175 | |
| GENERAL INFORMATION: | | | |
| APPLICANT: Penny, Laura | | | |
| APPLICANT: Galvin, Margaret | | | |
| APPLICANT: Miller, Andrew | | | |
| APPLICANT: Reidy, Michael | | | |
| TITLE OF INVENTION: Genotyping Human | | | |
| TITLE OF INVENTION: UDP-Glucuronosyltransferase 2B4 (UGT2B4), 2B7 (UGT2B7) and | | | |
| FILE REFERENCE: SEQ-22PRV2 | | | |
| CURRENT APPLICATION NUMBER: US/09/356,806 | | | |
| CURRENT FILING DATE: 1999-07-20 | | | |
| NUMBER OF SEQ. ID NOS: 164 | | | |
| SOFTWARE: FastSeq for Windows Version 3.0 | | | |
| SEQ ID NO: 8 | | | |
| LENGTH: 528 | | | |
| TYPE: PRT | | | |
| ORGANISM: H. sapiens | | | |
| US-09-356-806-8 | | | |
| Query Match | 93.0% | Score 2602.5 | DB 4; Length 528; |
| Best Local Similarity | 93.0% | Pred. No. 4.7e-263; | |
| Matches 492; | Conservative 14; | Mismatches 22; | Indels 1; Gaps 1; |
| QY | 1 | MSKMTSALLLIOLSCYFSSGCGKVLVPTFSHMNITITIDELVORGEVTVLASSA | 60 |
| DB | 1 | MSKMTSALLLIOLSCYFSSGCGKVLVPTFSHMNITITIDELVORGEVTVLASSA | 60 |
| QY | 61 | SISDPSPSTLKEFVYVSLTTEFEDIYKOLVKAELPKDTFMSYFSOVQIEIMTFN | 120 |
| DB | 61 | SISDPSPSTLKEFVYVSLTTEFEDIYKOLVKAELPKDTFMSYFSOVQIEIMTFN | 120 |
| QY | 121 | DIIRKFKDVSNNKMKLQESRFVDVLADAVPFEGELAEILKIDPVYSLRSPGYAI | 180 |
| DB | 121 | DIIRKFKDVSNNKMKLQESRFVDVLADAVPFEGELAEILKIDPVYSLRSPGYAI | 180 |
| QY | 181 | EKISGGLFPSPYVPMVMSLSDMTIERVKNIYLYEFWQITDMKKWQFYSBVL | 240 |
| DB | 181 | EKISGGLFPSPYVPMVMSLSDMTIERVKNIYLYEFWQITDMKKWQFYSBVL | 240 |
| QY | 241 | GRPTTSETMAKADIMLRVYMPDFPHLLPVEFVGGLHCKRPKPLPKMEEFVSSG | 300 |
| DB | 241 | GRPTTSETMAKADIMLRVYMPDFPHLLPVEFVGGLHCKRPKPLPKMEEFVSSG | 300 |
| QY | 301 | ENGWVPSLSQWVSNTESEERANYIASALAKIPOKVLWRFQGNKPDITGLNTRLYKWIPON | 360 |
| DB | 301 | ENGWVPSLSQWVSNTESEERANYIASALAKIPOKVLWRFQGNKPDITGLNTRLYKWIPON | 360 |

Query Match 83.0%; Score 2322; DB 4; Length 524;
Best Local Similarity 81.5%; Pred. No. 9.7e-234; Indels 0; Gaps 0;
Matches 427; Conservative 42; Mismatches 55;

US-09-356-806-40

US-09-356-806-40

Sequence 40, Application US/09356806

Patent No. 6586175

GENERAL INFORMATION:

APPLICANT: Penny, Laura

APPLICANT: Galvin, Margaret

APPLICANT: Miller, Andrew

APPLICANT: Reidy, Michael

TITLE OF INVENTION: Genotyping Human

TITLE OF INVENTION: UDP-Glucuronosyltransferase 2B4 (UGT2B4), 2B7 (UGT2B7) and

FILE REFERENCE: SEQ-22PRV2

CURRENT APPLICATION NUMBER: US/09/356,806

CURRENT FILING DATE: 1999-07-20

NUMBER OF SEQ ID NOS: 164

SOFTWARE: FastSeq for Windows Version 3.0

SEQ ID NO 40

LENGTH: 524

TYPE: PRT

ORGANISM: H. sapiens

US-09-356-806-40

Query Match 83.0%; Score 2322; DB 4; Length 524;
Best Local Similarity 81.5%; Pred. No. 9.7e-234; Indels 0; Gaps 0;
Matches 427; Conservative 42; Mismatches 55;

US-09-356-806-40

US-09-356-806-40

Sequence 40, Application US/09356806

Patent No. 6586175

GENERAL INFORMATION:

APPLICANT: Penny, Laura

APPLICANT: Galvin, Margaret

APPLICANT: Miller, Andrew

APPLICANT: Reidy, Michael

TITLE OF INVENTION: Genotyping Human

TITLE OF INVENTION: UDP-Glucuronosyltransferase 2B4 (UGT2B4), 2B7 (UGT2B7) and

FILE REFERENCE: SEQ-22PRV2

CURRENT APPLICATION NUMBER: US/09/356,806

CURRENT FILING DATE: 1999-07-20

NUMBER OF SEQ ID NOS: 164

SOFTWARE: FastSeq for Windows Version 3.0

SEQ ID NO 40

LENGTH: 524

TYPE: PRT

ORGANISM: H. sapiens

US-09-356-806-40

Query Match 77.5%; Score 2169.5; DB 3; Length 530;
Best Local Similarity 74.7%; Pred. No. 8.6e-218; Indels 1; Gaps 1;
Matches 396; Conservative 60; Mismatches 73;

US-09-180-852-2

US-09-180-852-2

Sequence 2, Application US/09180852

Patent No. 6287834

GENERAL INFORMATION:

APPLICANT: BELANGER, Alain

APPLICANT: HUM, Dean W.

APPLICANT: BEAULIEU, Martin

APPLICANT: LEVESQUE, Eric

TITLE OF INVENTION: CHARACTERIZATION AND USE OF AN ISOLATED URIDINE

TITLE OF INVENTION: DIHYDROPHOSPHO-GLUCURONOSYLTRANSFERASE

FILE REFERENCE: 1259-449

CURRENT APPLICATION NUMBER: US/09/180,852

CURRENT FILING DATE: 1999-02-08

EARLIER APPLICATION NUMBER: PCT/CA97/00328

EARLIER FILING DATE: 1997-05-16

EARLIER APPLICATION NUMBER: US 08/649,319

NUMBER OF SEQ ID NOS: 2

SOFTWARE: PatentIn Ver. 2.0

SEQ ID NO 2

LENGTH: 530

TYPE: PRT

ORGANISM: Homo sapiens

US-09-180-852-2

Query Match 77.5%; Score 2169.5; DB 3; Length 530;
Best Local Similarity 74.7%; Pred. No. 8.6e-218; Indels 1; Gaps 1;
Matches 396; Conservative 60; Mismatches 73;

US-09-180-852-2

US-09-180-852-2

Sequence 2, Application US/09180852

Patent No. 6287834

GENERAL INFORMATION:

APPLICANT: BELANGER, Alain

APPLICANT: HUM, Dean W.

APPLICANT: BEAULIEU, Martin

APPLICANT: LEVESQUE, Eric

TITLE OF INVENTION: CHARACTERIZATION AND USE OF AN ISOLATED URIDINE

TITLE OF INVENTION: DIHYDROPHOSPHO-GLUCURONOSYLTRANSFERASE

FILE REFERENCE: 1259-449

CURRENT APPLICATION NUMBER: US/09/180,852

CURRENT FILING DATE: 1999-02-08

EARLIER APPLICATION NUMBER: PCT/CA97/00328

EARLIER FILING DATE: 1997-05-16

EARLIER APPLICATION NUMBER: US 08/649,319

NUMBER OF SEQ ID NOS: 2

SOFTWARE: PatentIn Ver. 2.0

SEQ ID NO 2

LENGTH: 530

TYPE: PRT

ORGANISM: Homo sapiens

US-09-180-852-2


```

US-09-356-806-113
; Sequence 113, Application US/09356806
; Patent No. 6586175
; GENERAL INFORMATION:
; APPLICANT: Penny, Laura
; APPLICANT: Galvin, Margaret
; APPLICANT: Miller, Andrew
; APPLICANT: Reidy, Michael
; TITLE OF INVENTION: Genotyping Human
; TITLE OF INVENTION: UDP-glucuronosyltransferase 2B4 (UGT2B4), 2B7 (UGT2B7) and
; FILE REFERENCE: SEQ ID NO 2
; CURRENT APPLICATION NUMBER: US/09/356,806
; CURRENT FILING DATE: 1999-07-20
; NUMBER OF SEQ ID NOS: 164
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 113
; LENGTH: 530
; TYPE: PRT
; ORGANISM: H. sapiens
US-09-356-806-113

Query Match      77.1%; Score 2158.5; DB 4; Length 530;
Best Local Similarity 74.5%; Pred. No. 1.2e-216;
Matches 395; Conservative 60; Mismatches 74; Indels 1; Gaps 1;

QY 1 MSKMTSALLLIQSLSCYFSSGCGKVLVMPFESHMMNITITIDELVORGHETVLAASA 60
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1 MSLKMTSVFLIIQSLSCYFSSGCGKVLVMPFESHMMNITITIDELVORGHETVLAASA 60

QY 61 SISFDPSPETLKEEYVPSLTKTEFEDIIKOLVKRNA-ELPKDTFMSYFQVOEIMWTF 119
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 61 STLVNAKSSAIKIEVYFTSLTKNDLSDLIKIDRIYIGSKTFTMSYFQDELCEY 120

QY 120 NDILRKECKDIVSNKKLMKKLQESRFDVVLADAVFPFGELLAELDKLPVYSLRSPGYA 179
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 121 YDVSNKLCKDAVLNKKLMKKLQESKFDVILADALNPGCELLAELEFNIPLVLSLAFSGY 180

QY 180 IEKSGGLLPSPSYVPMVMSLSQMTFIERVKMIVLYLVEFPOFIDMKMKQDFYSEV 239
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 181 FEKGGGFLPPSPYVPMVMSLSQMTFIERVKMIVLYLVEFPOFIDMKMKQDFYSEV 240

QY 240 LGRPTTSETMAKADIMLRNVMDFOPPHPLPNVEFVGLHCKPAKPLPKMEEFVQSS 299
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 241 LGRPTTSETMAKADIMLRNVMDFOPPHPLPNVEFVGLHCKPAKPLPKMEEFVQSS 300

QY 300 GENGIVFSLGSMVSNTESEERANYIASALAKIPOKVLWRPDKNRPDLGLNTRLKYKIPQ 359
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 301 GENGIVFSLGSMVSNTESEERANYIASALAKIPOKVLWRPDKNRPDLGLNTRLKYKIPQ 360

QY 360 NDILGHPKTAFTTHGNGNGIYEAIYHGVPMVGVPIFGDDQDNTAHMKAKAAVEINFKT 419
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 361 NDILGHPKTAFTTHGNGNGIYEAIYHGVPMVGVPIFGDDQDNTAHMKAKAAVEINFKT 420

QY 420 MTSDDLRLATVTITDSSYKENAMRLSRIHDDQVKKPLDRAVPMIEFVWRHKGAKHLRSA 479
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 421 MTSDDLRLATVTITDSSYKENAMRLSRIHDDQVKKPLDRAVPMIEFVWRHKGAKHLRSA 480

QY 480 AHDLTWFQHSIDVIGFLITCVATAIFLTKCFLFSCKPFNKTREKRE 529
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 481 AHDLTWFQHSIDVIGFLITCVATAIFLTKCFLFSCKPFNKTREKRE 530

RESULT 5
US-09-813-918-2
; Sequence 2, Application US/09813918
; Patent No. 6383789
; GENERAL INFORMATION:
; APPLICANT: WEBSTER, Marion et al.
; TITLE OF INVENTION: ISOLATED HUMAN DRUG-METABOLIZING
; TITLE OF INVENTION: PROTEINS, NUCLEIC ACID MOLECULES ENCODING HUMAN
; TITLE OF INVENTION: DRUG-METABOLIZING PROTEINS,
; AND USES THEREOF

```

```

; FILE REFERENCE: CL0001175
; CURRENT APPLICATION NUMBER: US/09/813,918
; CURRENT FILING DATE: 2001-03-22
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 454
; TYPE: PRT
; ORGANISM: Human
US-09-813-918-2

Query Match      67.0%; Score 1874.5; DB 4; Length 454;
Best Local Similarity 67.9%; Pred. No. 4.6e-187;
Matches 359; Conservative 35; Mismatches 60; Indels 75; Gaps 1;

QY 1 MSKMTSALLLIQSLSCYFSSGCGKVLVMPFESHMMNITITIDELVORGHETVLAASA 60
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1 MSLKMTSVFLIIHLSYFSSGCGKVLVMAEYSHMMNMTITILKELVORGHETVLAASA 60

QY 61 SISFDPSPETLKEEYVPSLTKTEFEDIIKOLVKRNAELPKDTFMSYFQVOEIMWTFN 120
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 61 SILFDPNDASTLKEEYVPSLTKTEFENITMOQVKRMSDKOSFWLYFQBOEILMELY 120

QY 121 DILRKECKDIVSNKKLMKKLQESRFDVVLADAVFPFGELLAELDKLPVYSLRSPGYAI 180
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 121 DIFRNPKDVSNNKVMKKLQELRPDIVADAVFPCELLAALLNI----- 166

QY 181 EKSGGLLPSPSYVPMVMSLSQMTFIERVKMIVLYLVEFPOFIDMKMKQDFYSEV 240
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 167 ----- 166

QY 241 GRPTTSETMAKADIMLRNVMDFOPPHPLPNVEFVGLHCKPAKPLPKMEEFVQSSG 300
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 167 -RPTTLEFTHGKADIMLRNPMSPFPPPLPNVDVFGHCHCKPAKPLPKMEEFVQSSG 225

QY 301 ENGIVFSLGSMVSNTESEERANYIASALAKIPOKVLWRPDKNRPDLGLNTRLKYKIPQ 360
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 226 ENGIVFSLGSMVSNTESEERANYIASALAKIPOKVLWRPDKNRPDLGLNTRLKYKIPQ 285

QY 361 DILGHPKTAFTTHGNGNGIYEAIYHGVPMVGVPIFGDDQDNTAHMKAKAAVEINFKT 420
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 286 DILGHPKTAFTTHGNGNGIYEAIYHGVPMVGVPIFGDDQDNTAHMKAKAAVEINFKT 345

QY 421 TSEDLLRLATVTITDSSYKENAMRLSRIHDDQVKKPLDRAVPMIEFVWRHKGAKHLRSA 480
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 346 TSEDLLRLATVTITDSSYKENAMRLSRIHDDQVKKPLDRAVPMIEFVWRHKGAKHLRSA 405

QY 481 HDLTWFQHSIDVIGFLITCVATAIFLTKCFLFSCKPFNKTREKRE 529
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 406 HDLTWFQHSIDVIGFLITCVATAIFLTKCFLFSCKPFNKTREKRE 454

RESULT 6
US-09-813-918-3
; Sequence 3, Application US/09813918
; Patent No. 6383789
; GENERAL INFORMATION:
; APPLICANT: WEBSTER, Marion et al.
; TITLE OF INVENTION: ISOLATED HUMAN DRUG-METABOLIZING
; TITLE OF INVENTION: PROTEINS, NUCLEIC ACID MOLECULES ENCODING HUMAN
; TITLE OF INVENTION: DRUG-METABOLIZING PROTEINS,
; AND USES THEREOF
; FILE REFERENCE: CL0001175
; CURRENT APPLICATION NUMBER: US/09/813,918
; CURRENT FILING DATE: 2001-03-22
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 288
; TYPE: PRT
; ORGANISM: Human
US-09-813-918-3

```

Query Match 44.4%; Score 1243; DB 4; Length 288;
Best Local Similarity 80.2%; Pred. No. 2,3e-121;
Matches 231; Conservative 21; Mismatches 36; Indels 0; Gaps 0;

Qy 242 RPTLSEMAVADILINNYWDFOPHPLPNVEFVGLCKPAKPLKEMEFPQSSGE 301
Db 1 RPTLFETMGADILMNSWSFOPHPLPNVDFVGFHCKPAKPLKEMEFPQSSGE 60
Qy 302 NGVVFSLGSMVSNSEERANVIASALAKIPQKVLWRDGNKPDGLGNTLYKMI POND 361
Db 61 NGVVFSLGSMVSNSEERANVIASALAKIPQKVLWRDGNKPDGLGNTLYKMI POND 120
Qy 362 LGHPKTAFTTHGGANGIYEAHYGVPMVGPVIFGDOLDNIAMHAKGAAVEINFKMT 421
Db 121 LGHPKTAFTTHGGANGIYEAHYGVPMVGPVIFGDOLDNIAMHAKGAAVEINFKMT 180
Qy 422 SEDLRALRTVITDSSYKENAMRLSRHHDDQVPLDPAVFIEMVHKKAKILRSAAH 481
Db 181 STDLMNAKTYINDLYKENIMKLSRIHQDPVKPLDPAVFIEMVHKKAKILRSAAH 240
Qy 482 DLTWFQHSIDVIGFLTCVATATFLTKCFLESCQKFNKTRKIEKE 529
Db 241 DLTWFQHSIDVIGFLTCVATATFLTKCFLESCQKFNKTRKIEKE 288

RESULT 7

PCT-US92-00282-6
Sequence 6, Application PC/TUS9200282

GENERAL INFORMATION:
APPLICANT: OWENS, IDA S.
APPLICANT: RITTER, JOSEPH K.
TITLE OF INVENTION: THE GENETIC LOCUS UGT1 AND A MUTATION
TITLE OF INVENTION: THEREIN.
NUMBER OF SEQUENCES: 40
CORRESPONDENCE ADDRESS:
ADDRESSEE: CUSHMAN DARBY & CUSHMAN
STREET: 1615 L STREET, N.W.
CITY: WASHINGTON
STATE: D.C.
COUNTRY: U.S.A.
ZIP: 20036-5601
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US92/00282
FILING DATE: 19920110
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: SCOTT, WATSON T.
REGISTRATION NUMBER: 26581
REFERENCE/DOCKET NUMBER: 91532-PCT
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-861-3000
TELEFAX: 202-822-0944
TELEX: 6714627 CUSH
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 531 amino acids
TYPE: AMINO ACID
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
PCT-US92-00282-6

Query Match 42.8%; Score 1196.5; DB 5; Length 531;
Best Local Similarity 45.8%; Pred. No. 4.4e-116;
Matches 241; Conservative 104; Mismatches 172; Indels 9; Gaps 7;
Qy 3 MKMTSALLIOLSCYFSSGSCGVVWPTFESHMNNITILDELVORGHEVTLASASAI 62
Db 1 MKMTSALLIOLSCYFSSGSCGVVWPTFESHMNNITILDELVORGHEVTLASASAI 62

Db 7 LRGLSGILLILCALPMAEG--GKVLVPMESGSHLSNRDVREIARHQAVLAPETV 64
Qy 63 SFDNSSTLKFEVYPPVSLTKTEBEDIIKQLVKRMALPDTMFSYSSQVEIMWTFNDI 122
Db 65 HMKGEDPFTIQTAF--YTKBEYREILGNKAGFE-PQHPVKTFEFTMASIK-KFFDL 120
Qy 123 LRKFCIDIVSNKKMLKLOSRPDVVLADAVFPFGELLAEILKIPYVSLRSPGAIK 182
Db 121 YANSCALLHNKTLIQOLNSSPDVVLTDVFPFGALLAKYLOIPAVFPLRSVP-CGIDY 179
Qy 183 HSGGLLPSPVYVPMSELSDQMTFIRVKNMIVLYFEFMFQJEDMKKDDQFSEVLGR 242
Db 180 EATCCPSSSYTIPMLTMSDPMTFLORVXNMLYPLTKYICHL-STTPYESLSELLOR 238
Qy 243 PTLSEMAVADILINNYWDFOPHPLPNVEFVGLCKPAKPLKEMEFPQSSGEN 302
Db 239 EMSLVEVLSHSAVLPFGDFPDPYPIPMVMTFGGINCYIKPLSOFEAYANASGEH 298
Qy 303 GVVFSLGSMVSNSEERANVIASALAKIPQKVLWRDGNKPDGLGNTLYKMI POND 362
Db 299 GIVFSLGSMVSEIPEKKAMEIALGRIPQTLIMRYTGRPSVLAKNTLLVKMLPOND 358
Qy 363 LGHPKTAFTTHGGANGIYEAHYGVPMVGPVIFGDOLDNIAMHAKGAAVEINFKMTS 422.
Db 359 LGHPKTAFTTHGGANGIYEGICNGVPMVMPVIFGDOLDNAKMETRGAGVTLVLMETA 418
Qy 423 EDLRALRTVITDSSYKENAMRLSRHHDDQVPLDPAVFIEMVHKKAKILRSAAH 482
Db 419 DDLENNAKTYINNSYKENIRLSLHKDRIEPLDPAVFIEMVHKKAKILRSAAH 478
Qy 483 LTFWFQHSIDVIGFLTCVATATFLTKCFLESCQKFNKTRKIEKE 527
Db 479 LTFWFQHSIDVIGFLTCVATATFLTKCFLESCQKFNKTRKIEKE 524

RESULT 8

PCT-US92-00282-3
Sequence 3, Application PC/TUS9200282

GENERAL INFORMATION:
APPLICANT: OWENS, IDA S.
APPLICANT: RITTER, JOSEPH K.
TITLE OF INVENTION: THE GENETIC LOCUS UGT1 AND A MUTATION
TITLE OF INVENTION: THEREIN.
NUMBER OF SEQUENCES: 40
CORRESPONDENCE ADDRESS:
ADDRESSEE: CUSHMAN DARBY & CUSHMAN
STREET: 1615 L STREET, N.W.
CITY: WASHINGTON
STATE: D.C.
COUNTRY: U.S.A.
ZIP: 20036-5601
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US92/00282
FILING DATE: 19920110
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: SCOTT, WATSON T.
REGISTRATION NUMBER: 26581
REFERENCE/DOCKET NUMBER: 91532-PCT
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-861-3000
TELEFAX: 202-822-0944
TELEX: 6714627 CUSH
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 533 amino acids
TYPE: AMINO ACID
STRANDEDNESS: single

```

;      TOPOLOGY: linear
;      MOLECULE TYPE: protein
PCT-US92-00282-3

```

| | | | | |
|---------------------------|--------|---------------------|------------|-------------|
| Query Match | 40.4%; | Score 1130.5; | DB 5; | length 533; |
| Best Local Similarity | 44.7%; | Pred. No. 3.5e-109; | | |
| Matches 232; Conservative | 89; | Mismatches 153; | Indels 35; | Gaps 7; |

```

QY 24 GKLYMPTESHHMMKIKTIDELVORGHVTVLASSASISFDPNSPSTLKEVPE----- 78
Db 28 GKTLILPVDSHMLSMUGAIQOOLQOGHGVTVLADPADSLYIRDOGAFTYLLK--TVPEPOR 85
QY 79 -----VSLTKTEFED--IIKOLVGRMAELPKDFWFSYFSOVOEIMWTFNDILRKECD 129
Db 86 EDVKESFVSLGHNVHFENDESFLORVIKTYKXIKKD-----SAMLSCSH 129
QY 130 IVSNKKLIMKKLOESRFVDVLADAVPEPGBELLAEILLKIPVYSLRSPGALIEKHSGLLF 189
Db 130 LLHNKEIMLMASSSFDVMLTDPFLPCSPIVAQYLSLPTVFLLHLP--CSLEFATQCPN 188
QY 190 PPSYVPVMSLSDDQMTFIERVKNMIVLYIYEFWFQIIDMKKMOFSEVLGRTTILSET 249
Db 189 PFSYVPRPLSSHSDHMTFLORVKNMLIARSONFLCDVY--YSRYATLASEFLOREVTODL 247
QY 250 MAQADIMLRNYMDFQPHPLLBNVEFVGGLHCKRAXLLPREMEBFVSGSENVFSL 309
Db 248 LSSASVWLFPSDVKQYPRPIMPNNVFGVGINCLHQNPISQEPFAYINASEHGIVFSL 307
QY 310 GSNVSNSTSEERANVIASALAKIPOKLIMRFDCNKPDDTGLMTRLYKWIPOUNDLGHPRK 369
Db 308 GSNVSELPPEKKAMAILDALGKNPQYLYMRYTTRSNLANNTILVKMLPOUNDLGHMTR 367
QY 370 AFTTHGMMNGIYEAIHGVPMWGVPIFGDOLDNIAMHAKGAAYEINFKTWTSDDLRLAL 429
Db 368 AFTTHAGSHGVYESICNGVPMWMPPLFGQMDNARMEETKAGVTLINLEWTSDDLNAL 427
QY 430 RYVIYTSSTYKSNAMRLSRHHDDPYKPLDRAYFWLIEFVNRHKGAHLRSAAHDLTWQYH 489
Db 428 KAVINIKSYKENTMRSSLSHKORPVEPPLDLAFWVEFPMRHKGAHLRPAHDLTWQYH 487
QY 490 SIIVIGLILCVATAIFLFTKCLSPSCQK--FNKTKRIK 527
Db 488 SLDIVIGLLAVLVLTVAIFTPCCAGAYGCLGKGRVKK 526

```

```

1 RESULT 9
2 PCT-US92-00282-4
3
4 ; Sequence 4, Application PC/TUS9200282
5 ; GENERAL INFORMATION:
6 APPLICANT: OWENS, IDA S.
7 APPLICANT: RITTER, JOSEPH K.
8 TITLE OF INVENTION: THE GENETIC LOCUS UGT1 AND A MUTATION
9 TITLE OF INVENTION: THEREIN.
10 NUMBER OF SEQUENCES: 40
11 CORRESPONDENCE ADDRESS:
12 ADDRESSEE: CUSHMAN DARBY & CUSHMAN
13 STREET: 1615 L STREET, N.W.
14 CITY: WASHINGTON
15 STATE: D.C.
16 COUNTRY: U.S.A.
17 ZIP: 20036-5601
18
19 COMPUTER READABLE FORM:
20 MEDIUM TYPE: Floppy disk
21 COMPUTER: IBM PC compatible
22 OPERATING SYSTEM: PC-DOS/MS-DOS
23 SOFTWARE: PatentIn Release #1.0, Version #1.25
24
25 CURRENT APPLICATION DATA:
26 APPLICATION NUMBER: PCT/US92/00282
27 FILING DATE: 19920110
28 CLASSIFICATION: 435
29 ATTORNEY/AGENT INFORMATION:
30 NAME: SCOTT, WATSON T.
31 REGISTRATION NUMBER: 26581

```

```

? TELEPHONE/DOCKET NUMBER: 91532-PC7
?
? TELECOMMUNICATION INFORMATION:
?
? TELEPHONE: 202-861-5000
?
? TELEFAX: 202-822-0944
?
? TELEX: 6714627 CUSH
?
? INFORMATION FOR SEQ ID NO: 4:
?
? SEQUENCE CHARACTERISTICS:
?
? LENGTH: 534 amino acids
?
? TYPE: AMINO ACID
?
? STRANDEDNESS: single
?
? TOPOLOGY: linear
?
? MOLECULE TYPE: protein
?
? PCT-US92-00282-4

```

| | | | | |
|-----------------------|------------------|-------------------|------------|-------------|
| Query Match | 40.2%; | Score 1126; | DB 5; | Length 534; |
| Best Local Similarity | 43.9%; | Pred. No. 1e-108; | | |
| Matches 234; | Conservative 93; | Mismatches 176; | Indels 30; | Gaps 7 |

| | | | | |
|----|--|-----|---|-----|
| Oy | | 6 | TSALLIIQISCVSSGSCGVLMPTFESHMNIIITIDELVQGHETVLAASASISFD | 65 |
| Db | | 14 | TGILLILISQPMABE---GKVLVPTDQSPMLSNREALREIHAHQAVLITPEVNHIK | 70 |
| Oy | | 66 | PNSPTLKEEYVPSVLTKEPEDI-----IKOLVKMAELPKDTFWSYFSQVOEI | 115 |
| Db | | 71 | BEKEFTL--TAAVAPMTOKEPDRVLTIGVTOGFETBHLK-----YSRSMAI | 116 |
| Oy | | 116 | MMTFENDILKPKCDIVSNKKLKKLOESRFDVLTADAVFPGBELLABELKIPVYLSRFS | 175 |
| Db | | 117 | MNVNSIALRCCEVLEIHNELIRHNANSPDVLITDPNLICGAVLAKYLSIPAFPMRYI | 176 |
| Oy | | 176 | PGVAIKESHGGLFPSPYVPVMSLSQMTFIEKVMKIYVLXFEFMFOIPDMKKKDOF | 235 |
| Db | | 177 | P-CDLDFKQTCQPNBSYIPKLLTNSHMTFLRKVMKMLPLALSYICHTFS-APVAST | 234 |
| Oy | | 236 | YSEVLGRPTLSETMAKADIMLIRNYPMDFOPHPLPVEVFGJHCCKAPKLPKEMEER | 295 |
| Db | | 235 | ASLIFREYVSVDLYSYASVMLFRDPMYDIPRPLPMVMVITGGINANGKPLSGEEAY | 294 |
| Oy | | 296 | VQSGENGUVVSLGSMVSNTESEERANVIAAALAKIPQVLMRPDGNKPTLGINTRLYK | 355 |
| Db | | 295 | INASGSHGIVFSLGSMSEIPEKKAMAIADALGKITPTVLMRTGTRPSILANITLYK | 354 |
| Oy | | 356 | WIFQNDLICHPKTAPATHGGMNGIYEAIYHGVPWGVPIFGDOLONIAHMKAGAAVEI | 415 |
| Db | | 355 | WLPQNDLLGHMPTRAFATHAGSHGYEISICGVPPVMMPRLFGDOMDAKMETEKGAGVTL | 414 |
| Oy | | 416 | NFTMTSEBLLALAKTVITDSSYKKNAMRLSRIHNDQVYKRLDRAVPMIEFMVNHKAKH | 475 |
| Db | | 415 | NVLEMTSEBLENALAKVINDKSYKENIRLSSLHNDREVEBLDAVPMVEFMVNHKAGPH | 474 |
| Oy | | 476 | LRSAAHDLMPQHSIDVYGFLLTCVATAAIFLFTCFLSCOK--FNKTRKLEK | 527 |
| Db | | 475 | LRPAAHDLMPQHSIDVYGFLLAVLVTAFTFCKCAVYGRKLCGKKGRVYK | 527 |

```

1 RESULT 10
2 PCT-US92-00282-5
3 Sequence 5, Application PC/TUS9200282
4 GENERAL INFORMATION:
5 APPLICANT: OWENS, IDA S
6 APPLICANT: RITTER, JOSEPH K.
7 TITLE OF INVENTION: THE GENETIC LOCUS UGT1 AND A MUTATION
8 TITLE OF INVENTION: THEREIN.
9 NUMBER OF SEQUENCES: 40
10 CORRESPONDENCE ADDRESS:
11 ADDRESSEE: CUSHMAN DABY & CUSHMAN
12 STREET: 1615 L STREET, N.W.
13 CITY: WASHINGTON
14 STATE: D.C.
15 COUNTRY: U.S.A.
16 ZIP: 20036-5601
17 COMPUTER READABLE FORM:
18 MEDIUM TYPE: Floppy disk
19

```

```

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US92/00282
FILING DATE: 19920110
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: SCOTT, WATSON T.
REGISTRATION NUMBER: 26581
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-861-3000
TELEFAX: 202-822-0944
TELEX: 6714627 CUSH
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 531 amino acids
TYPE: AMINO ACID
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
PCT-US92-00282-5

Query Match      38.0%; Score 1064; DB 5; Length 531;
Best Local Similarity 43.4%; Pred. No. 3.1e-102;
Matches 221; Conservative 90; Mismatches 180; Indels 18; Gaps 8;

25 KVLVWPTFSSHMNIKITLDELVORGHEVTVLASSASISFDPNSPTLKFEVYVSLTKT 84
28 KLLVVPQSGSHWLSMKDIVEVLSDRGHEIVVVPVNNLLKRYKTYRK--IYVPVYDQ 85
85 EFEDIKOLVYKMAELPKDTFWS-YFSQVQELMTFNDILRKF-----CKDIVSNKKMK 139
86 E-----LKKRYGSGFNHFAERSFLTAPQTEYNNMIVGLYFNCQLQDQDNTNF 138
140 LQESRFDVLADAVFPFGLLAELKIPVYSLRSPGVALEKHSGLLFPSPVYVPM 199
139 FKESRFDALFTDPALPCGVIIAELVGLSVYLFRRFP--CSLEHTSRSPDPVSYIPRCYT 197
200 ELSDQMTFIERVKMIYVLYFEFWFOIFDMKKMDQFSEVILGRPTLTSETMAKADIWLR 259
198 KFSDHMTFSQRYANLVNLIBEYLFYCL-FSKYEKLAGAVLKRVDII-TLSEVSVMLLR 255
260 NYWDQFPHPLLPNVEFVGLHCKRAKPLPKEMEEFVQSSGNGVNVVSLGSMVNTSEE 319
256 YDFVLEYRPPVMPNVPFGLGNCCKRKQDSQEFVAVINASGEHGLVSVLSGWSVEIPEK 315
320 RANVASALAKIPQKVLWRFDPGNKPDTLGLNTRLYKMI PONDLLGHPTKAFITTHGMNG 379
316 KAMALADALGNPQVULMKRYGTRPSNLANNITLVKMLPONDLLGHPTKAFITTHAGSHG 375
360 IYEAIVHGVNVPVPIFGDQDLDNIAMKAGAAVEINFKMTSDDLRLALRTVITDSSYK 439
376 VYESICNGVPMVMDLFGDQMDNAKMETKGAGVTLNVLMTSEDLNALKAVINDKSYK 435
440 ENAMLSRIHDPKPLDRAVPMIEFVWRHKGAGHLSAANDLTWFOHYSIDVIGPLLT 499
436 ENIMLSLSLHKDRPVEPLDLAVFWVEFVWRHKGAGHLPRAADLTWYOHSLDIVIGFLLA 495
500 CVATAIPLFTKCFLFSCCK-FNKTRKIEK 527
496 VVLTAIPLFTKCFPFYGRKCGKGRVKK 524

```

```

NUMBER OF SEQUENCES: 40
CORRESPONDENCE ADDRESS:
ADDRESSEE: CUSHMAN DABRY & CUSHMAN
STREET: 1615 L STREET, N.W.
CITY: WASHINGTON
STATE: D.C.
COUNTRY: U.S.A.
ZIP: 20036-5601
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US92/00282
FILING DATE: 19920110
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: SCOTT, WATSON T.
REGISTRATION NUMBER: 26581
REFERENCE/DOCKET NUMBER: 91532-PCT
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-861-3000
TELEFAX: 202-822-0944
TELEX: 6714627 CUSH
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 529 amino acids
TYPE: AMINO ACID
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
PCT-US92-00282-7

Query Match      37.4%; Score 1045.5; DB 5; Length 529;
Best Local Similarity 41.1%; Pred. No. 2.7e-100;
Matches 211; Conservative 101; Mismatches 174; Indels 27; Gaps 7;

25 KVLVWPTFSSHMNIKITLDELVORGHEVTVLASSASISFDPNSPTLKFEVYVSLTKT 84
27 KLLVVPQSGSHWLSMKDIVEVLSDRGHEIVVVPVNNLLKESKRYRKSPVYNNL--- 83
85 EFEDIKOLVYKMAELPKDTFWSYFSQVQELMTFNDILRKF-----CKDIVSNKK 135
84 -----ELRRRYNSFGN---NHFASSPLMAPLREYNNMIVYDMCFSCQSLKQSA 133
136 LMKLQESRFPVLADAVFPFGLLAELKIPVYSLRSPGVALEKHSGLLFPSPVYVPM 195
134 TLFSLRENQFALFTDPAMPQGVIIAELYKLPISYLFRRGP--CSLEHIGSPSPVSYV 191
196 VVMSLSQMTFIERVKMIYVLYFEFWFOIFDMKKMDQFSEVILGRPTLTSETMAKADI 255
192 RPYTKFSHMTFPPRLAFLINIENTLYHCL-YSKTEILASDLKQDVL--PALHONS 249
256 WLIRNYMDFOPPHPLLPNVEFVGLHCKRAKPLPKEMEEFVQSSGNGVNVVSLGSMVSN 315
250 WLLAYDFEYRPPVMPNVPFGLGNCCKRKQDSQEFVAVYNASGEHGLVSVLSGWSVE 309
316 TSEERANVASALAKIPQKVLWRFDPGNKPDTLGLNTRLYKMI PONDLLGHPTKAFITTHG 375
310 IPEKKAMIBALGRIPQTLIMRYTGTRPSNLAQNTLVKMLPONDLLGHPTKAFITTHG 369
376 GMNGIYEAIVHGVNVPVPIFGDQDLDNIAMKAGAAVEINFKMTSDDLRLALRTVITD 435
370 GSHGIYEGICNGVPMVMDLFGDQMDNAKMETKGAGVTLNVLMTADLNLAKTVANN 429
436 SSKYENAMLSRIHDPKPLDRAVPMIEFVWRHKGAGHLSAANDLTWFOHYSIDVIG 495
430 KSYKENIMRSLSLHKDRPVEPLDLAVFWVEFVWRHKGAGHLPRAADLTWYOHSLDIVIG 489
496 FLITCVATAIPLFTKCFLFSCCK-FNKTRKIEK 527
490 FLTAIPLTVFVIYKSCAYGCRKCFGKGRVKK 522

```

RESULT 12

US-09-813-918-4
Sequence 4, Application US/09813918
Patent No. 6383789
GENERAL INFORMATION:
APPLICANT: WEBSTER, Marion et al.
TITLE OF INVENTION: ISOLATED HUMAN DRUG-METABOLIZING
TITLE OF INVENTION: PROTEINS, NUCLEIC ACID MOLECULES ENCODING HUMAN
TITLE OF INVENTION: DRUG-METABOLIZING PROTEINS,
FILE REFERENCE: CL001175
CURRENT APPLICATION NUMBER: US/09/813,918
CURRENT FILING DATE: 2001-03-22
NUMBER OF SEQ ID NOS: 4
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 4
LENGTH: 197
TYPE: PRT
ORGANISM: Human
US-09-813-918-4

Query Match 29.8%; Score 833; DB 4; Length 197;
Best Local Similarity 78.2%; Pred. No. 8.8e-79;
Matches 154; Conservative 18; Mismatches 25; Indels 0; Gaps 0;

QY 1 MSMWTSAALLIQLSCYSSGCGKVLVWPTFESHMNIKITLDELVORGHEVTLASSA 60
DB 1 MTLKMTSVLLIHLISCYFSSGCGKVLVMAEYSHMMNMKITLDELVORGHEVTLASSA 60
QY 61 SISDPNPSPTLKEFVVPVSLTKTEFEDIKQLVKRWAELEPKDTFMSYFSQVGEIMMTFN 120
DB 61 SILFDPNDASTLKEVPTSLTKTEFENILMQVKRMSDIKDSFWLYFSOEILLWELY 120
QY 121 DILKRFCDIVSNKMKKLQESRPDVVLADAVFPFGLLAELIKIPVYSLRFSPGYAI 180
DB 121 DIFNFCDDVSNKMKKLQESRPDIYFADAVFPFGLLAELINIRVYSLRFTPGYTI 180
QY 181 EKHSGLFPSPYV 197
DB 181 ERHSGGLFPSPYIPV 197

RESULT 13

US-09-305-856B-18
Sequence 18, Application US/09305856B
Patent No. 6479236
GENERAL INFORMATION:
APPLICANT: Penny, Laura
TITLE OF INVENTION: Genotyping the Human
TITLE OF INVENTION: UDP-Glucuronosyltransferase 1 (UGT1) Gene
FILE REFERENCE: 4389-7 (formerly SEQ-17CIP)
CURRENT APPLICATION NUMBER: US/09/305,856B
CURRENT FILING DATE: 1999-05-05
PRIOR APPLICATION NUMBER: 60/084,807
PRIOR FILING DATE: 1998-05-07
NUMBER OF SEQ ID NOS: 124
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 18
LENGTH: 245
TYPE: PRT
ORGANISM: Homo sapiens
US-09-305-856B-18

Query Match 27.4%; Score 765.5; DB 4; Length 245;
Best Local Similarity 60.1%; Pred. No. 1.4e-71;
Matches 143; Conservative 35; Mismatches 59; Indels 1; Gaps 1;
QY 291 EMEFFVQSGGNGVYVSLGSMVNTSERANVIALAKIPKYLWRFQGNKPDITGLN 350
DB 1 EFAFYINASBGHGIIVFSLGSMVSEIPEKKAMIALDALGKIPIQTVLWRYTGTBRSNLANN 60

QY 351 TRLYKWIPONDLLGHPKTKAFITTHGNGGIYEAIYHGVPMVGVPIFGDOLDNIAHMKAKG 410
DB 61 TILVKMLPQNDLLGHPKTRAFITTHAGSHGYEISCNVPMVMPPLFDDQDNNAKMKETKG 120
QY 411 AAVEINKMTSDDLALRTVITDSSYKENAMRLSIHHDQPKPLDRAVWPIEFVWRH 470
DB 121 AGTIANVLEMTSDDLALRTVITDSSYKENAMRLSIHHDQPKPLDRAVWPIEFVWRH 180
QY 471 KGAKHLRSANADLTWPFQHSYDVIQPLTCAVAIRLFTFCFLFSCCK-FNKRRIKIK 527
DB 181 KGAPHLRPAADLTWYQYHSLDVIQPLAVLTVAFITFCACAYGRKCLGKGRVAK 238

RESULT 14

US-08-942-012B-32
Sequence 32, Application US/08942012B
Patent No. 6235278
GENERAL INFORMATION:
APPLICANT: Lu, Albert
APPLICANT: Miller, Lois K.
APPLICANT: Dietze, Peter
APPLICANT: Black, Bruce
TITLE OF INVENTION: Biological Insect Control Agents Expressing
TITLE OF INVENTION: Insect-Specific Toxin Genes, Methods and Compositions
FILE REFERENCE: 28-96a
CURRENT APPLICATION NUMBER: US/08/942,012B
CURRENT FILING DATE: 1997-10-01
PRIOR APPLICATION NUMBER: 08/729,606
PRIOR FILING DATE: 2000-10-01
NUMBER OF SEQ ID NOS: 33
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 32
LENGTH: 515
TYPE: PRT
ORGANISM: Spodoptera littoralis nuclear polyhedrosis virus
US-08-942-012B-32

Query Match 14.4%; Score 403; DB 3; Length 515;
Best Local Similarity 26.7%; Pred. No. 3.7e-33;
Matches 141; Conservative 102; Mismatches 227; Indels 58; Gaps 20;

QY 9 LLLIQLSCYSSGCGKVLVWPT-FESHMNIKITLDELVORGHEVTLASSASISDPN 67
DB 5 ILVVSILVLRNSAAVRLCMFTPTPSYSHQVPPVYVYVALLRKHSLVIT--SKPIHNNH 62
QY 68 SPSTLKEFVVPVSLTKTEFEDIKQLVKRWAELEPKDTFMSYFSQVGEIMMTFNDILR--- 124
DB 63 GHRHHRHE---NLTEIDVGSVTNNPFKR---LLQD---SKVSRKGIIVSDSTVTRVNY 112
QY 125 -----KFCDDIVSNKMKKLQESRPDVVLADAVFPFGLLAELIKIPVYSLRFS 175
DB 113 LGLARMISAQFEHGV--KRLRSNQ--TPDVIVIEAFVSPILSYFFKDTPTV--IQIS 166
QY 176 PGVAIEKH---SGGLLEPPSYVPMVMSSELDQMTFIERVGNMI--VYLVEFPWFQIPD-- 228
DB 167 SGHGTAEINFTMGAVARHPVYPMWRDRKGLSWQVTAQVTEIRLHYEFSQLDADQS 226
QY 229 --MKWDQFYSEVILGREPTLSETMAKADILIRNYWDFOPPHLPLNVEFVGLHCKP-- 284
DB 227 AMMKR--QFGSKV---PDVDALRKVYHMFVNTHPVFTNRRPVPSPVQVIGIHDPAV 280
QY 285 ---AKPLPKMEBFVQSGGNGVYVSLGSMV--SNTSERANVIALAKIPKYLWRF 339
DB 281 TSVADEIDNDIAEFLENS-TMGVYVYVSLGSSVASDMSNMLVFEFTPSIPRYLVAKV 339
QY 340 DGNKP---DTLGLNTRLYKWIPONDLLGHPKTKAFITTHGNGGIYEAIYHGVPMVGVPIFG 397
DB 340 DKSDKFDNIPSNVLIQRFQRRVLRKRVKVFITQSGVQSDAIDAGVPMFQVPIMG 399
QY 398 DOLDNIAHMKAKAAVEINKMTSDDLALRTVITDSSYKENAMRLSIHHDQPKPL 457
DB 400 DQFYNYVWYETYGIGRCVDTLTVDARQLTEIVMDVADNEKXKKGITLWLRDAINDQPMRPL 459

Oy 458 DRAVWIEFVNRHKGK-KHLRSAAHDLTFWFOHYSID-VIGFLLTCVAT 503
Db 460 EKAVWYTHVARAKGKHGLTRAAVYTSKYAMPDLILPMLITFST 507

RESULT 15

US-09-370-838-36
; Sequence 36, Application US/09370838
; Patent No. 6444425
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Mohamath, Roadoh
; APPLICANT: Secrist, Heather
; TITLE OF INVENTION: COMPOUNDS FOR THERAPY AND DIAGNOSIS OF
; FILE REFERENCE: 210121.475C1
; CURRENT APPLICATION NUMBER: US/09/370,838
; CURRENT FILING DATE: 1999-08-09
; EARLIER APPLICATION NUMBER: US 09/285,323
; EARLIER FILING DATE: 1999-04-02
; NUMBER OF SEQ ID NOS: 289
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 36
; LENGTH: 129
; TYPE: PRT
; ORGANISM: Homo sapien
US-09-370-838-36

Query Match 14.0%; Score 393; DB 4; Length 129;
Best Local Similarity 58.1%; Pred. No. 4.3e-33;
Matches 75; Conservative 19; Mismatches 35; Indels 0; Gaps 0;

Oy 303 GVVVFSLSGMSVSNISEERANVIASALAKIPQKVLNRFPGNKPDTLGLNTRLYKWIPONDL 362
Db 1 GVVVFSLSGMSVSEIPEKKAVALADALGKIPQTVLWRYGTGTRPSNLANNITLWQWLPONDL 60
Oy 363 LGHPKTKAFITHGNGNGIYEATYHGVPMVGVPIFGDOLDNIAHMKAGAAVEINFKMTS 422
Db 61 LGHPKTRAFITTHASGHVNESICNGVPMVPIPLFGDQMDNAKRRETKAGAVTLNVLEMTS 120
Oy 423 EDLRLRLRT 431
Db 121 EDLEDALKS 129

Search completed: December 5, 2003, 09:50:50
Job time : 22 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 7, 2003, 03:53:47 ; Search time 4104 Seconds

(without alignments)
16447.584 Million cell updates/sec

Title: US-09-980-729B-10

Perfect score: 1650

Sequence: 1 agcaactcggaagaacagcat.....ccaattcaagaagacctg 1650

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 2888711 seqs, 2045481386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : GenEmbl:
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rtd:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sv:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|--------|-------------|--------|--------------|---------------------|
| 1 | 1650 | 100.0 | 1650 | 6 AX040087 | AX040087 Sequence |
| 2 | 1444.4 | 87.5 | 2079 | 9 HSAJ5162 | AX005162 Homo sapi |
| 3 | 1443.8 | 87.5 | 2129 | 9 BC026264 | BC026264 Homo sapi |
| 4 | 1433.2 | 86.9 | 2097 | 9 AF064200 | AF064200 Homo sapi |
| 5 | 1432.6 | 86.8 | 2093 | 6 AX411198 | AX411198 Sequence |
| 6 | 1432.6 | 86.8 | 2033 | 9 HSDUPGT | Y00317 Human mRNA |
| 7 | 1431.6 | 86.8 | 2091 | 9 AF081793 | AF081793 Homo sapi |
| 8 | 1331.4 | 80.7 | 1639 | 6 AX548042 | AX548042 Sequence |
| 9 | 1292.4 | 78.3 | 2108 | 9 AF112112 | AF112112 Macaca fa |
| 10 | 1267.8 | 76.8 | 1766 | 9 BC030974 | BC030974 Homo sapi |
| 11 | 1254.6 | 76.0 | 1915 | 9 AF401657 | AF401657 Macaca fa |
| 12 | 1252.6 | 75.9 | 1855 | 6 AX336329 | AX336329 Sequence |
| 13 | 1252.6 | 75.9 | 1855 | 6 AX336696 | AX336696 Sequence |
| 14 | 1252.6 | 75.9 | 1855 | 6 AX409473 | AX409473 Sequence |
| 15 | 1252.6 | 75.9 | 1855 | 9 HUMDPEGTA | U05428 Human 3',4-c |
| 16 | 1248 | 75.6 | 2648 | 9 MF091582 | U01582 Macaca fa |
| 17 | 1238.6 | 75.1 | 1753 | 9 AF016310 | AF016310 Macaca fa |
| 18 | 1236.8 | 75.0 | 1722 | 9 AF016492 | AF016492 Homo sapi |
| 19 | 1236.8 | 75.0 | 1768 | 9 AF112113 | AF112113 Macaca fa |
| 20 | 1234.4 | 74.8 | 1662 | 9 AF294902 | AF294902 Macaca mu |
| 21 | 1232.8 | 74.7 | 1662 | 9 AF294901 | AF294901 Macaca mu |
| 22 | 1225.2 | 74.3 | 2799 | 6 AX411112 | AX411112 Sequence |
| 23 | 1225.2 | 74.3 | 2799 | 9 HSDUGT2BIO | X63359 H.sapiens U |
| 24 | 1208.6 | 73.2 | 1851 | 9 AF177272 | AF177272 Homo sapi |
| 25 | 1159.6 | 70.3 | 2107 | 6 AR168316 | AR168316 Sequence |
| 26 | 1159.6 | 70.3 | 2107 | 9 HSUS8209 | US9209 Homo sapien |
| 27 | 1145.2 | 69.4 | 2114 | 9 AF180322 | AF180322 Homo sapi |
| 28 | 1133 | 68.7 | 2092 | 9 AF548389 | AF548389 Homo sapi |
| 29 | 1125.2 | 68.2 | 2090 | 6 AX410646 | AX410646 Sequence |
| 30 | 1125.2 | 68.2 | 2090 | 9 HSU08854 | U08854 Human UDP g |
| 31 | 1115 | 67.6 | 2123 | 6 AX410640 | AX410640 Sequence |
| 32 | 1115 | 67.6 | 2123 | 9 HSU06641 | U06641 Human UDP g |
| 33 | 1110.4 | 67.3 | 2075 | 9 AF072223 | AF072223 Macaca fa |
| 34 | 1100.2 | 60.7 | 1671 | 4 RABUGT2B | LO1082 Oryctolagus |
| 35 | 967.4 | 58.6 | 1636 | 6 AX327326 | AX327326 Sequence |
| 36 | 967.4 | 58.6 | 2759 | 6 AX395164 | AX395164 Sequence |
| 37 | 967.4 | 58.6 | 2974 | 6 AX376454 | AX376454 Sequence |
| 38 | 965.8 | 58.5 | 3006 | 6 AX497163 | AX497163 Sequence |
| 39 | 962.6 | 58.3 | 2818 | 4 OCUT7242 | U2742 Oryctolagus |
| 40 | 962.2 | 58.3 | 1832 | 4 RABUGT2BA | LO1081 Oryctolagus |
| 41 | 956.8 | 58.0 | 1961 | 6 AX401727 | AX401727 Sequence |
| 42 | 956.8 | 58.0 | 1961 | 10 RATUDPCTP | M13506 Rat liver U |
| 43 | 955 | 57.9 | 2560 | 10 BC027200 | BC027200 Mus muscu |
| 44 | 947.4 | 57.4 | 1825 | 10 AB034987 | AB034987 Cavia por |
| 45 | 936.8 | 56.8 | 1584 | 6 AX497165 | AX497165 Sequence |

ALIGNMENTS

RESULT 1
AX040087
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL

AX040087 1650 bp DNA linear PAT 18-NOV-2000
Sequence 10 from Patent WO0063351.
AX040087.1 GI:11230049
Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
Lal,P., Yue,H., Tang,Y.T., Hillman,J.L., Baughn,M.R. and Yang,J.
Carbohydrate-modifying enzymes
Patent: WO 0063351-A 10 26-OCT-2000;

Incyte Genomics, Inc. (US)

Location/Qualifiers

1. 1650

/organism="Homo sapiens"

/mol_type="genomic DNA"

/db_xref="taxon:9606"

/note="Incyte ID No: 2912330CB1"

BASE COUNT 489 a 330 c 354 g 477 t

ORIGIN

Query Match 100.0%; Score 1650; DB 6; Length 1650;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 1650; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 AGCACTGGAAGAAACAGAGATTCGATCGATGAGATGCTATGAAATGAGCTGAGCT 60
 1 AGCACTGGAAGAAACAGAGATTCGATCGATGAGATGCTATGAAATGAGCTGAGCT 60
 61 TCTCTGATACAGCTGAGCTGTTACTTTAGCTCTGGAGTTGTGGAAGTCTGTGTG 120
 61 TCTCTGATACAGCTGAGCTGTTACTTTAGCTCTGGAGTTGTGGAAGTCTGTGTG 120
 121 GCCCAGAGATTCAGCCACTGATGATATTAAGCAATCTCGATGAATCTTGTCCAG 180
 121 GCCCAGAGATTCAGCCACTGATGATATTAAGCAATCTCGATGAATCTTGTCCAG 180
 181 AGGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 240
 181 AGGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 240
 241 ATCTACTGTTAAATTTGAAGTTATCTGATCTTTAACTAAACTGATTTGAGATAT 300
 241 ATCTACTGTTAAATTTGAAGTTATCTGATCTTTAACTAAACTGATTTGAGATAT 300
 301 TATCAGAGAGCTGTTAAGAGATGAGAGATGAGAGATGAGAGATGAGAGATGAG 360
 301 TATCAGAGAGCTGTTAAGAGATGAGAGATGAGAGATGAGAGATGAGAGATGAG 360
 361 TTCAAGAGATGAGAGATGAGAGATGAGAGATGAGAGATGAGAGATGAGAGATGAG 420
 361 TTCAAGAGATGAGAGATGAGAGATGAGAGATGAGAGATGAGAGATGAGAGATGAG 420
 421 TATAGTTTCAATAAGAACTTATGAGAACTAGAGAGTCAAGATTTGATGTTCT 480
 421 TATAGTTTCAATAAGAACTTATGAGAACTAGAGAGTCAAGATTTGATGTTCT 480
 481 TGCAGATGCTGTTTCCCTTTGTGAGCTGCTGAGCCGAGATTCTTAAATACCTTTGT 540
 481 TGCAGATGCTGTTTCCCTTTGTGAGCTGCTGAGCCGAGATTCTTAAATACCTTTGT 540
 541 CTACAGCTCCGCTCTCTCTGCTGAGCAATTAAGAACTAGTGAAGCACTTCTGT 600
 541 CTACAGCTCCGCTCTCTCTGCTGAGCAATTAAGAACTAGTGAAGCACTTCTGT 600
 601 CCTCTCTCTATGCTGCTGTTATGTCAGAACTAGTGAAGCACTTCTCATAGA 660
 601 CCTCTCTCTATGCTGCTGTTATGTCAGAACTAGTGAAGCACTTCTCATAGA 660
 661 GAGGGTAAATAATGATCTATGCTTATTTGAATTTGGTCCAAATATTTGACAT 720
 661 GAGGGTAAATAATGATCTATGCTTATTTGAATTTGGTCCAAATATTTGACAT 720
 721 GAGGAAGTGGATCTAGTCTACAGTGAAGTTTGAAGAACTAGTGAAGCACTTCTGAG 780
 721 GAGGAAGTGGATCTAGTCTACAGTGAAGTTTGAAGAACTAGTGAAGCACTTCTGAG 780
 781 AATGCAAAAGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 840
 781 AATGCAAAAGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 840
 841 ACTCTTACCAATGTTGAGTTGTTGAGAGCTCCAGTCAAACTGCAAACTGCAAACT 900
 841 ACTCTTACCAATGTTGAGTTGTTGAGAGCTCCAGTCAAACTGCAAACTGCAAACT 900

901 GAAGAAATGGAAGATTGTCAGAGCTGAGAAATGTTGTTGTTCTCT 960
 901 GAAGAAATGGAAGATTGTCAGAGCTGAGAAATGTTGTTGTTCTCT 960
 961 GGGGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1020
 961 GGGGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1020
 1021 CAAGATCCCAAAAGTTCTGTGAGATTTGATGGAATTAACGATTTTATGAGAT 1080
 1021 CAAGATCCCAAAAGTTCTGTGAGATTTGATGGAATTAACGATTTTATGAGAT 1080
 1081 CAATCTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1140
 1081 CAATCTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1140
 1141 AGCTTTTATCACTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1200
 1141 AGCTTTTATCACTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1200
 1201 TATGTTGAGATTTCCATTTTGTGATGATGATGATGATGATGATGATGATGATGAT 1260
 1201 TATGTTGAGATTTCCATTTTGTGATGATGATGATGATGATGATGATGATGATGAT 1260
 1261 AGGAGAGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1320
 1261 AGGAGAGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1320
 1321 GAGAAAGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1380
 1321 GAGAAAGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1380
 1381 CCATGATCACTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1440
 1381 CCATGATCACTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1440
 1441 CCAGAAAG 1500
 1441 CCAGAAAG 1500
 1501 CTCTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1560
 1501 CTCTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1560
 1561 AAAATGTTTATTTTCTGTCAGAAATTTAAATTAAGAACTAGTGAAGCACTTCTG 1620
 1561 AAAATGTTTATTTTCTGTCAGAAATTTAAATTAAGAACTAGTGAAGCACTTCTG 1620
 1621 ATAGATCTTTCCAAATCAAGAAAGACCTG 1650
 1621 ATAGATCTTTCCAAATCAAGAAAGACCTG 1650

RESULT 2
 HSAJ5162
 DEFINITION Homo sapiens mRNA for UDP-glucuronosyltransferase
 ACCESSION AJ005162.1 GI:3135024
 VERSION UDP-glucuronosyltransferase; UGT2B4 gene.
 KEYWORDS
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 Jackson, M.R., McCarthy, L.R., Harding, D., Wilson, S., Coughtrie, M.W.
 and Burchell, B.
 Cloning of a human liver microsomal UDP-glucuronosyltransferase
 cDNA
 JOURNAL Biochem. J. 242 (2), 581-588 (1987)
 MEDLINE 87241362
 PUBMED 3109396

REFERENCE 2
AUTHORS Rittner, J. K., Chen, F., Sheen, Y. Y., Lubet, R. A., and Owens, I. S.
TITLE Two human liver cDNAs encode UDP-glucuronosyltransferases with 2
log differences in activity toward parallel substrates including
hydroxycholeic acid and certain estrogen derivatives
JOURNAL Biochemistry 31 (13), 3409-3414 (1992)
MEDLINE 92207964
PUBMED 1554722
REFERENCE 3 (bases 1 to 2079)
AUTHORS Rittner, J. K.
TITLE Direct Submission
JOURNAL Submitted (08-MAY-1998) Rittner J. K., Pharmacology and Toxicology,
Virginia Commonwealth University, P.O. Box 980613, Medical College
of Virginia, Richmond, Virginia, 23112-0613, USA
FEATURES
SOURCE location/Qualifiers
1..2079
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="UDPGTb-1"
/clone_lib="Lambda bacteriophage gt11"
/tissue_lib="Liver"
1..2079
/gene="UGT2B4"
34..1620
/gene="UGT2B4"
/EC_number="2.4.1.17"
/function="transfers glucuronic acid to hydroxycholeic
acid and estrogen catechols"
/codon_start=1
/evidence=experimental
/product="UDP-glucuronosyltransferase"
/protein_id="CA06396.1"
/db_xref="GI:3135025"
/db_xref="SPTRMBL:O60867"
/translation="MSMKTSALLLIQLSCYSSGCGKLVLPTEFSHMNNIKIILD
ELVORGHVTVLASSASISFDNPSPTLKFVYVPSLTKTEFEDIKOLVRMELPK
DTFWSYSONOEIFMTNDILRKCKDIVSKKIMKLOESRPVVLADAVPFGELL
AELIKIPYVSLRSPGAIIRKSGGLFPSPSYFVPMSELSDDPTIRKMYITVL
YFEFPOIFDMKKWDQFSEVLGRPTLSETMAADLILNWDFOFPHLPLNVEF
VGLHCKRPAKPLPEMEBFVQSGENGVSLSMVSNTSSEERANYIASLAKI PK
VLNRFDGNKPTGLNTRLYKMI PONDILGHPKTRAFITGGANGIYEAIVHGIPMG
VPLPADONDIAHMKAKGAVALDFHTWSSTDLNALAKTVINDPLUYENAMKLSRIH
DQPKPLDRAVFWIEFVWRHKGAKHLRAVAHDLTWFQHSLSLDTVTFLLACATVIFII
TKCLFCYWKVFRTRGKGRD"
546
/gene="UGT2B4"
/citation=[1]
/replace="g"
548..549
/gene="UGT2B4"
/citation=[1]
/replace="ct"
1187..1188
/gene="UGT2B4"
/citation=[1]
/replace="ct"
1190
/gene="UGT2B4"
/citation=[1]
/replace="c"
1192
/gene="UGT2B4"
/citation=[1]
/replace="a"
1977..1978
/gene="UGT2B4"
/citation=[1]
/replace="cga"
2068
/gene="UGT2B4"
/citation=[1]
/replace="g"

BASE COUNT 634 a 396 c 436 g 613 t
ORIGIN
Query Match 87.5%; Score 1444.4; DB 9; Length 2079;
Best Local Similarity 93.3%; Pred. No. 0;
Matches 1521; Conservative 0; Mismatches 106; Indels 3; Gaps 1;
QY 2 GCAACTGGAAAAACAAGATGTCATTCAGATGTCATGAATGAGACTTCAGCTCTT 61
DB 1 GCAACTGGAAAAACAAGATGTCATTCAGATGTCATGAATGAGACTTCAGCTCTT 60
QY 62 CTCCTGATACAGCTGAGCTGTCTTACTTTAGCTCTGGAGTTGTGAAAGTGTGCTG 121
DB 61 CTGCTGATACAGCTGAGCTGTCTTACTTTAGCTCTGGAGTTGTGAAAGTGTGCTG 120
QY 122 CCCACGAATTCAGCCCTGAGTGAATATTAAGCAATCTCTGATGTAATCTTCCAGAGA 181
DB 121 CCCACGAATTCAGCCCTGAGTGAATATTAAGCAATCTCTGATGTAATCTTCCAGAGA 180
QY 182 GGTGATGAGGTAAGTATGATGGCATCTTCAGCTTCCATTTCTTTCGATCCCAAGCCCA 241
DB 181 GGTGATGAGGTAAGTATGATGGCATCTTCAGCTTCCATTTCTTTCGATCCCAAGCCCA 240
QY 242 TCTACTCTTAATTTGAAGTTATCTGTATCTTTAACTTAACTGAGTTGAGGATATT 301
DB 241 TCTACTCTTAATTTGAAGTTATCTGTATCTTTAACTTAACTGAGTTGAGGATATT 300
QY 302 ATCAAGAGCTGTTTAAGAGATGGGCGAACTTCCAAAAGACATTTTGGTCAATTTT 361
DB 301 ATCAAGAGCTGTTTAAGAGATGGGCGAACTTCCAAAAGACATTTTGGTCAATTTT 360
QY 362 TCACAAATACAGAAATCATGTGACATTTAATGACATCTAGTAAAGTTGTGTAAGAT 421
DB 361 TCACAAATACAGAAATCATGTGACATTTAATGACATCTAGTAAAGTTGTGTAAGAT 420
QY 422 ATAGTTCAATAAGAAACTATGTAAGAACTACAGAGATCAAGATTTGATGTTCTT 481
DB 421 ATAGTTCAATAAGAAACTATGTAAGAACTACAGAGATCAAGATTTGATGTTCTT 480
QY 482 GCAGATGCTGTTTCCCTTTCCTTGGTGAAGTGTGCGGCGAGTTACTTTAAATACCTTTGTC 541
DB 481 GCAGATGCTGTTTCCCTTTCCTTGGTGAAGTGTGCGGCGAGTTACTTTAAATACCTTTGTC 540
QY 542 TAGAGCTCGGCTCTGCTCGGCTAGCAATTGAAAGCATAGTGGAGACTTCTGTTTC 601
DB 541 TAGAGCTCGGCTCTGCTCGGCTAGCAATTGAAAGCATAGTGGAGACTTCTGTTTC 600
QY 602 CCTCCTTCTATGCTGCTGTTGTTATGTGCAAGATCAAGATCACTTTGATAGAG 661
DB 601 CCTCCTTCTATGCTGCTGTTGTTATGTGCAAGATCAAGATCACTTTGATAGAG 660
QY 662 AGGGTAAAAAATATGATGATGCTTATTTTGAATTTTGGTCCAAATTTTGAACATG 721
DB 661 AGGGTAAAAAATATGATGATGCTTATTTTGAATTTTGGTCCAAATTTTGAACATG 720
QY 722 AAGAAGTGGATGATGCTCTACAGTGAAGTTCTAGAGAACCCACTAGCTTATGAGACA 781
DB 721 AAGAAGTGGATGATGCTCTACAGTGAAGTTCTAGAGAACCCACTAGCTTATGAGACA 780
QY 782 ATGGCAAAAGCTGACATATGCTTATTTGAAACTGAGATTTTCAATTTCTCACCCA 841
DB 781 ATGGCAAAAGCTGACATATGCTTATTTGAAACTGAGATTTTCAATTTCTCACCCA 840
QY 842 CTCTTACCAAAATGTTGATGCTGTTGAGAGACTCCACTGCAAACTGCCAAACCCCTACCG 901
DB 841 CTCTTACCAAAATGTTGATGCTGTTGAGAGACTCCACTGCAAACTGCCAAACCCCTACCG 900
QY 902 AAGGAATGGAAGATTTGTCAGAGCTGAGGAAATGAGTGTGTTGTTTCTCTG 961
DB 901 AAGGAATGGAAGATTTGTCAGAGCTGAGGAAATGAGTGTGTTGTTTCTCTG 960
QY 962 GGGTCATGTCATGTAACAGCTCAGAAAGAAAGGCAATGTAATGATCAGCCCTTGGC 1021
DB 962 GGGTCATGTCATGTAACAGCTCAGAAAGAAAGGCAATGTAATGATCAGCCCTTGGC 1021

Center, Stanford University School of Medicine, Stanford, CA 94305
Web site: <http://www.sngc.stanford.edu>
Contact: (Dickson, Mark) mcd@paxil.stanford.edu
Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers, R. M.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: <http://image.lnl.gov>
Series: IRAL Plate: 37 Row: C Column: 2
This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 10863940.

FEATURES
source
1. .2129
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="LocusID:7363"
/db_xref="taxon:9606"
/clone="MGC:22623 IMAGE:4709436"
/issue_type="Liver"
/clone_id="NH_MGC_76"
/lab_host="DH10B"
/note="Vector: pDNR-LIB"
45. .1631
/codon_start=1
/product="UDP glycosyltransferase 2 family, polypeptide B4"
/protein_id="AAH26264.1"
/db_xref="gi:20070776"
/translation="MSMKTSALLLIQLSCYSSGSGKVLVPEFSHMNIKITILL
ELVORGEVAVLASASISPDNPSPTLKEFVPSLTKFEDIKOLVRMAELPPELID
DTPFSFNOVETIMMTFNDILRKRCIOIVSKXLMKLOESRPVLADAVPFGELIL
AELIKIPVYSLRFSPGAIIEKSGGLFPFSYIPVVMSELSDMTPIERKXNIIYV
YEEFWFOIFDMKRWDOFYSEVLRPTLSEMAVADWLINWDFQPSHPLPVEE
VGGHCKPAPKLPKEMEEFVSGSENGVSELSMVSNTSEERANYALAKI PQE
VLMPDGNKPEPTLGLNRLYKWI PONDILGPKTRAPITFGANGIYEAIVHQIPMW
VPELPADPNIAHMKGAAGAAVSLDFHMTSSDILNALKTVINDLPYENAKLSIHF
DQPKPLDRAVFWIEFVMRHKGAGHLVAHADLWFPQHSLDVGFLLACATATYIFI
TKCLFVWKFRGKGRD"

BASE COUNT 672 a 398 c 441 g 618 t
ORIGIN
Query Match 87.5% Score 1443.8; DB 9; Length 2129;
Best Local Similarity 93.3% Pred. No.0; Mismatches 107; Indels 3; Gaps 1
Matches 1521; Conservative 0;

1 AGCAACTGGAAGAAAGCATTGCATTCAGATGTCATGAATGACTTCAGCTCT 60
11 AGCAACTGGAAGAAAGCATTGCATTCAGATGTCATGAATGACTTCAGCTCT 70
61 TCTCCTGATACAGCTGAGCTGTACTTAACTTACCTGCGAGTTGTGGAAGTGCCTGTG 120
71 TCTGCTGATACAGCTGAGCTGTACTTAACTTACCTGCGAGTTGTGGAAGTGCCTGTG 130
121 GCCCAAGAAATTCAGCCACGTGAAGAAATTAAGACATCCGTGATGAACCTTGCCAGAG 180
131 GCCCAAGAAATTCAGCCACGTGAAGAAATTAAGACATCCGTGATGAACCTTGCCAGAG 190
181 AGTCATGAGCTGACTGATTGGGACTTCCAGCTTCCATTTCTTTCGATCCCAACAGCCC 240
191 AGTCATGAGCTGACTGATTGGGACTTCCAGCTTCCATTTCTTTCGATCCCAACAGCCC 250
241 ATCTACTCTTAAATTTGAAGTTTATCCTGTATCTTTAACTTTAACTGAAGCTTGAGAGAT 300
251 ATCTACTCTTAAATTTGAAGTTTATCCTGTATCTTTAACTTTAACTGAAGCTTGAGAGAT 310
301 TATCAAGAGCTGTTTAAGAGATGGGCAAGCTTCCAAAGACACATTTTGGTCATATTT 360
311 TATCAAGAGCTGTTTAAGAGATGGGCAAGCTTCCAAAGACACATTTTGGTCATATTT 370
361 TTCAAGATACAGAAATCATGTGGACATTTAATGACATCTTAGAAAGTTCTGTAAAGA 420
371 TTCAAGATACAGAAATCATGTGGACATTTAATGACATCTTAGAAAGTTCTGTAAAGA 430

```

Qy 421 TATAGTTCAATAGAGAACTTATGAGAGAACTACAGAGTCAAGATTGATGTTGTTCT 480
Db 431 TATAGTTCAATAGAGAACTTATGAGAGAACTACAGAGTCAAGATTGATGTTGTTCT 490
Qy 481 TGCAGATGCTGTTTCCCTTTGGTGGAGCTGCTGCGAGTACTTAAATACCTTTGT 540
Db 491 TGCAGATGCTGTTTCCCTTTGGTGGAGCTGCTGCGAGTACTTAAATACCTTTGT 550
Qy 541 CTACAGCTCCGCTTCTCTCCGCTACGCAATTGAAGAAGATGAGAGACTTCTGTT 600
Db 551 CTACAGCTCCGCTTCTCTCCGCTACGCAATTGAAGAAGATGAGAGACTTCTGTT 610
Qy 601 CCTCTCTCTATGCTGCTGTTGATGTCAGAACTAGAGCAATGACTTTATAGA 660
Db 611 CCTCTCTCTATGCTGCTGTTGATGTCAGAACTAGAGCAATGACTTTATAGA 670
Qy 661 GAGGTAATAAATATGATCTATGCTTATTTGATTTGTTGTTCCAAATATTTGACAT 720
Db 671 GAGGTAATAAATATGATCTATGCTTATTTGATTTGTTGTTCCAAATATTTGACAT 730
Qy 721 GAAGAAGTGGATGATCTACAGTGAAGTCTAGGAAGACCCACTAGCTATCGAGAC 780
Db 731 GAAGAAGTGGATGATCTACAGTGAAGTCTAGGAAGACCCACTAGCTATCGAGAC 790
Qy 781 AATGCAAAAGCTGACATATGCTTATTCGAACCTACTGAGATTTTCAATTTCTCAACC 840
Db 791 AATGCAAAAGCTGACATATGCTTATTCGAACCTACTGAGATTTTCAATTTCTCAACC 850
Qy 841 ACTCTTACCAATGTTGAGTTGCTGTTGAGAGACTCCACTGCAACCTGCAAAACCCCTAAC 900
Db 851 ACTCTTACCAATGTTGAGTTGCTGTTGAGAGACTCCACTGCAAAACCCCTAAC 910
Qy 901 GAAGAATGAGAGTGTTCAGAGCTCGAGAAATGCTGTGTTGTTGTTTCTCT 960
Db 911 GAAGAATGAGAGTGTTCAGAGCTCGAGAAATGCTGTGTTGTTGTTTCTCT 970
Qy 961 GGGGTGATGCTGATGACATGACAGTCAAGAAAGGCGCAATGATGATGATGATGATGATGAT 1020
Db 971 GGGGTGATGCTGATGACATGACAGTCAAGAAAGGCGCAATGATGATGATGATGATGATGAT 1030
Qy 1021 CAAGATCCCAAAAGGTTCTGTGAGATTTGATGGAATTAACCAATCTTTAGAGCT 1080
Db 1031 CAAGATCCCAAAAGGTTCTGTGAGATTTGATGGAATTAACCAATCTTTAGAGCT 1090
Qy 1081 CAATACTGCGCTGATCAAGTGGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1140
Db 1091 CAATACTGCGCTGATCAAGTGGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1150
Qy 1141 AGCTTTTATCACTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1200
Db 1151 AGCTTTTATCACTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1210
Qy 1201 TATGTTGGAGTTCCTCATATTTTGGTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1260
Db 1211 TATGTTGGAGTTCCTCATATTTTGGTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1270
Qy 1261 AGGAGCGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1320
Db 1271 AGGAGCGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1330
Qy 1321 GAGAAGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1380
Db 1331 GAGAAGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1390
Qy 1381 CGATGATCAACCTGTAAGCCCTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1440
Db 1391 CGATGATCAACCTGTAAGCCCTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1450
Qy 1441 CCAGAAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1500
Db 1451 CCAGAAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1510
Qy 1501 CTCTATAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1560

```

```

Db 1511 CTCTTGGAGTGAAGTGGTTCCTGCTGCGCTGTGGCAATGCTGATATTCATCATCAC 1570
Qy 1561 AAAATGTTTATTTATTTTCCGTCAAAATTTAATAAATAGAAAGATAGAAAGAGGA 1620
Db 1571 AAAATGTTTATTTTGTGTG--GAAGTTGTTAGAACAGAAAGAGGAGGAGGAGGAGGAGGAGGAG 1627
Qy 1621 ATAGATCTTC 1631
Db 1628 TTAATTACGTC 1638

RESULT 4
AF064200 2097 bp mRNA linear PRI 11-DEC-1998
LOCUS AF064200
DEFINITION Homo sapiens UDP-glucuronosyltransferase 2B4 precursor (UGT2B4)
ACCESSION AF064200
VERSION AF064200
KEYWORDS AF064200.1 GI:3153831
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE
AUTHORS Levesque, E., Beaulieu, M., Hum, M. and Belanger, A.
TITLE Characterization and substrate specificity of UGT2B4 (E458): A
JOURNAL UDP-glucuronosyltransferase encoded by a polymorphic gene
REFERENCE PharmacoGenetics (1999) In press
AUTHORS Levesque, E., Beaulieu, M. and Belanger, A.
JOURNAL Direct Submission
JOURNML Submitted (08-MAY-1998) Molecular Endocrinology, CHUL Research
FEATRES Center, 2705 Laurier Blvd, Ste-Foy, PQ G1V 4G2, Canada
source Location/Qualifiers
1..2097
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
1..2097
/gene="UGT2B4"
/allele="UGT2B4*E458"
26..1612
/gene="UGT2B4"
/allele="UGT2B4*E458"
/feature="E458 variant"
/codon_start=1
/product="UDP-glucuronosyltransferase 2B4 precursor"
/protein_id="AAC95002.1"
/db_xref="GI:3153832"
/translacion="MSMKWTSALLLIQLSCYFSGSGCKVLMPTFESHMWNITKIID
ELYORGEHVIVLSSAISFPDPSRSLTKREYVPSLTKEFEFIIQOLVKRAELPK
DIFWYSFVQOEIIMWTENDILRKCKOIVSNKIMKRLQSRPVLADAVFPFGEIL
AEILKIPFVYSLRPSGVALERKSGGLPFPSPVVPVMSLSLQDMEIEVKNNIYVL
YFERFQIPDMKMDQFYSLEVLGRPTLSLTKAKADIMLRNWDFOFPHPLPNVFR
VGGHLCRPAKPLKREMEFVOSGENGVVSLGSVNSSEERANYIASALAKIPOK
VLMRFDNRKPDITGLNRLYKWI PONDILGHPKTRAITGGAAGIYEALYHGIPIVG
VPLPADPNDIIMAKKAGAVSLDPRHMSSTDILNAIKTVINPLVYENMKVSRIRH
DQVKPLRERAVNIIEFMRHKAGHLEVAHDLTFWPHSLDVTGFLACVATYIFII
TKCLPCWKRVRVFGKGRD"
BASE COUNT 653 a 393 c 435 g 616 t
ORIGIN
Query Match 86.9%; Score 1433.2; DB 9; Length 2097;
Best Local Similarity 93.2%; Pred. No. 0;
Matches 1511; Conservative 0; Mismatches 108; Indels 3; Gaps 1;
Qy 10 AAAACAAGCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 69
Db 1 AAAACAAGCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 60
Qy 70 ACAAGTAGCTGTTACTTATGCTTGGAGTGTGGAAGGTGCTGTGGCCACAGA 129
Db 61 ACAAGTAGCTGTTACTTATGCTTGGAGTGTGGAAGGTGCTGTGGCCACAGA 120

```

```

Oy 130 ATTGAGCCATGATGATATTAAGACAATCCTGGATGAATCTGTCCAGAGAGGTCAAGA 189
Db 121 ATTCAGCCATGATGATATTAAGACAATCCTGGATGAATCTGTCCAGAGAGGTCAAGA 180
Oy 190 GGTGACTGATTTGGCATCTTCAGCTTCATTTCTTTCATCCCAACAGCCCATCTACTCT 249
Db 181 GGTGACTGATTTGGCATCTTCAGCTTCATTTCTTTCATCCCAACAGCCCATCTACTCT 240
Oy 250 TAAATTTGAAGTTTATCTGTATCTTTAACTTAACTGAGTTTGAGATTTATTAACGA 309
Db 241 TAAATTTGAAGTTTATCTGTATCTTTAACTTAACTGAGTTTGAGATTTATTAACGA 300
Oy 310 GCTGGTTAAGATGGGAGAACTTCCAAAGACATTTGGTCATATTTTTCACAAGT 369
Db 301 GCTGGTTAAGATGGGAGAACTTCCAAAGACATTTGGTCATATTTTTCACAAGT 360
Oy 370 ACAAGAAATCATGTGAGATTTTATGACATATTAAGAAATTTCTGTAAGATATAGTTTC 429
Db 361 ACAAGAAATCATGTGAGATTTTATGACATATTAAGAAATTTCTGTAAGATATAGTTTC 420
Oy 430 AAATAGAACTTATGAAAGAACTACAGAGTCAAGATTTGATGTTTCTTTCAGATGC 489
Db 421 AAATAGAACTTATGAAAGAACTACAGAGTCAAGATTTGATGTTTCTTTCAGATGC 480
Oy 490 TGTTCCTCCCTTGGAGCTGCGGAGTACTTAAATATACCTTTGTCTACAGGCT 549
Db 481 TGTTCCTCCCTTGGAGCTGCGGAGTACTTAAATATACCTTTGTCTACAGGCT 540
Oy 550 CCGCTTCTCTCCCTGCTACGCAATTTGAAAGCATAGTGGAGGACTTCTGTCCTCTTC 609
Db 541 CCGCTTCTCTCCCTGCTACGCAATTTGAAAGCATAGTGGAGGACTTCTGTCCTCTTC 600
Oy 610 CTATGCTCCTGTTGTTATGTCAAGAACTAAGTACCAATGACTTTCATAGAGAGGTAA 669
Db 601 CTATGCTCCTGTTGTTATGTCAAGAACTAAGTACCAATGACTTTCATAGAGAGGTAA 660
Oy 670 AAATAGATCTATGCTTATTTGAAATTTGGTCCAAATATTTGACATGAAAGTG 729
Db 661 AAATAGATCTATGCTTATTTGAAATTTGGTCCAAATATTTGACATGAAAGTG 720
Oy 730 GGATGATTTCTACAGTGAAGTTCTAGGAAGCACTACGTTATCTGAGACATGGCAA 789
Db 721 GGATGATTTCTACAGTGAAGTTCTAGGAAGCACTACGTTATCTGAGACATGGCAA 780
Oy 790 AGCTGACATATGCTTATTTGAAATCTACTGGATTTTCAATTTCTCACCCACTTTACC 849
Db 781 AGCTGACATATGCTTATTTGAAATCTACTGGATTTTCAATTTCTCACCCACTTTACC 840
Oy 850 AAATGTTAGTTGCTTGGAGGAGCTCACTGCAAACTGCAAACTGCAAACTGCAAACTG 909
Db 841 AAATGTTAGTTGCTTGGAGGAGCTCACTGCAAACTGCAAACTGCAAACTGCAAACTG 900
Oy 910 GGAAGAGTTTGTCCAGAGCTCTGAGAAATGATGTTGTGTTTCTCTGGGGTCAAT 969
Db 901 GGAAGAGTTTGTCCAGAGCTCTGAGAAATGATGTTGTGTTTCTCTGGGGTCAAT 960
Oy 970 GGTCTGTAACAGCTGAGAAAGGCGCAATGTAATTTGATCAAGCTTTCGCAAGTCCC 1029
Db 961 GGTCTGTAACAGCTGAGAAAGGCGCAATGTAATTTGATCAAGCTTTCGCAAGTCCC 1020
Oy 1030 ACAAAAGGTTCTGTGAGATTTGATGGGAAATTAACAGATCTTTTGAAGCTCAATATCG 1089
Db 1021 ACAAAAGGTTCTGTGAGATTTGATGGGAAATTAACAGATCTTTTGAAGCTCAATATCG 1080
Oy 1090 GCTGTACAAGTGGATACCCAGAAATGATCTTCTGGTCAATCCCAAAACCAAGCTTTTAT 1149
Db 1081 GCTGTACAAGTGGATACCCAGAAATGATCTTCTGGTCAATCCCAAAACCAAGCTTTTAT 1140
Oy 1150 CACTCATGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1209
Db 1141 AACTCATGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1200

```

```

Oy 1210 AGTTCATATTTTGGATGATCAGTTGATTAACATAGCTCACATGAGGCCAAAGAGCAGC 1269
Db 1201 CGTTCATATTTTGGATGATCAGTTGATTAACATAGCTCACATGAGGCCAAAGAGCAGC 1260
Oy 1270 TGTGAAATTAACCTTCAAACTATGACAGAGGAAATTTACTGAGGCTTTGAGAACGT 1329
Db 1261 TGTGAGTTTGGACCTTCAACAAATGTCAGAGACTACTCTCAATGCACTGAAGACGT 1320
Oy 1330 CATTACGATTCCTTATTAAGAAATGCTATAGATTTATCAAGAAATTTACCATGATCA 1389
Db 1321 AATTATGATCTTTATTAAGAAATGCTATAGATTTATCAAGAAATTTATCATGATCA 1380
Oy 1390 ACTGTAAAGCCCTAGATCAGAGCTTCTGATGATGATGATGATGATGATGATGATG 1449
Db 1381 ACCAGTGAAGCCCTTGAAGAGCAGCTTCTGATGATGATGATGATGATGATGATGATG 1440
Oy 1450 AGCCAGACCTGCGATCAGCTGCTCCATGACCTCACTGTTCCAGACTTCTATAGA 1509
Db 1441 AGCCAGACCTGCGATCAGCTGCTCCATGACCTCACTGTTCCAGACTTCTATAGA 1500
Oy 1510 TGTGATTTGGCTTCTGCTGACCTGTCGAGCACTGCTATATCTTCTTCAAAATGTT 1569
Db 1501 TGTGATTTGGCTTCTGCTGACCTGTCGAGCACTGCTATATCTTCTTCAAAATGTT 1560
Oy 1570 TTTATTTTCTGTCGCAAAATTTAATAAACTAGAAAGTAAAGAGGAAATGATCTT 1629
Db 1561 GTTTTGTCTG---GAAGTTTGTTAAGACAGGAAAGAGGAAAGATTAATTACG 1617
Oy 1630 TC 1631
Db 1618 TC 1619

```

RESULT 5

AX411198

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

TITLES

JOURNAL

FEATURES

SOURCE

BASE COUNT

ORIGIN

Query Match

Best Local Similarity

Matches 1514, Conservative

Score 1432.6; DB 6; Length 2093;

Pred. No. 0;

Mismatches 114; Indels 3; Gaps 1;

1 AGCAATGGAAGAAACAGATTCATGATGATGATGATGATGATGATGATGATGATGATG 60

4 AGCAATGGAAGAAACAGATTCATGATGATGATGATGATGATGATGATGATGATGATG 63

61 TCTCTGATACAGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 120

64 TCTGCTGATACAGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 123

121 GCCCAGAGATTCAGCCACTGGATGAATATTAAGCAATCTGATGAATCTTGTCCAGAG 180

124 GCCCAGAGATTCAGCCACTGGATGAATATTAAGCAATCTGATGAATCTTGTCCAGAG 183

QY 181 AGGATAGAGTGAAGTATGATGATCTGAGTCTTCCATTTCTTGCATCCCAAGAGCC 240
 DB 184 AGGATAGAGTGAAGTATGATGATCTGAGTCTTCCATTTCTTGCATCCCAAGAGCC 243
 QY 241 ATCTACTCTTAAATTTGAAGTTATCTGATCTTAACTAACTGAGTTTGAGATAT 300
 DB 244 ATCTACTCTTAAATTTGAAGTTATCTGATCTTAACTAACTGAGTTTGAGATAT 303
 QY 301 TATTAACAGCTGGTTAAGATGGGCAAACTTCCAAAGACATTTTGTGATATTT 360
 DB 304 TATTAACAGCTGGTTAAGATGGGCAAACTTCCAAAGACATTTTGTGATATTT 363
 QY 361 TTCAAGATCAAGAAATCATGATGAGATTTATGATCACTTGAAGTTCTGTAAGA 420
 DB 364 TTCAAGATCAAGAAATCATGATGAGATTTATGATCACTTGAAGTTCTGTAAGA 423
 QY 421 TATAGTTTCAAAATGAAGAACTTATGAAGAACTACAGAGTCAAGATTTGATGTTCT 480
 DB 424 TATAGTTTCAAAATGAAGAACTTATGAAGAACTACAGAGTCAAGATTTGATGTTCT 483
 QY 481 TGCAGATGCTGTTTCCCTTTGGTGAAGTCTGCGAGTCTTAAATACCTTTGT 540
 DB 484 TGCAGATGCTGTTTCCCTTTGGTGAAGTCTGCGAGTCTTAAATACCTTTGT 543
 QY 541 CTACAGCTCGCTTCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 600
 DB 544 CTACAGCTCGCTTCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 603
 QY 601 CCTCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 660
 DB 604 CCTCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 663
 QY 661 GAGGTTAAATATGATCTATGCTGCTTATTTGAATTTGGTCCAAATTTGATCAT 720
 DB 664 GAGGTTAAATATGATCTATGCTGCTTATTTGAATTTGGTCCAAATTTGATCAT 723
 QY 721 GAAGAAATGGGATAGTTCTACAGTGAAGTCTGGAAGAACCTGCTGCTGCTGCTGCT 780
 DB 724 GAAGAAATGGGATAGTTCTACAGTGAAGTCTGGAAGAACCTGCTGCTGCTGCTGCT 783
 QY 781 AATGGCAAAAGCTGATGATGCTTATTTGAAGTCTGGAATTTGATTTCTGCTGCT 840
 DB 784 AATGGCAAAAGCTGATGATGCTTATTTGAAGTCTGGAATTTGATTTCTGCTGCT 843
 QY 841 ACTCTTACCAATTTGATGCTTGAAGTCTGGAAGAACCTGCTGCTGCTGCTGCT 900
 DB 844 ACTCTTACCAATTTGATGCTTGAAGTCTGGAAGAACCTGCTGCTGCTGCTGCT 903
 QY 901 GAAGAAATGGGATAGTTCTACAGTGAAGTCTGGAAGAACCTGCTGCTGCTGCTGCT 960
 DB 904 GAAGAAATGGGATAGTTCTACAGTGAAGTCTGGAAGAACCTGCTGCTGCTGCTGCT 963
 QY 961 GGGGTGATGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1020
 DB 964 GGGGTGATGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1023
 QY 1021 CAAGATCCCAAAAGTTCTGCTGAGATTTGATGGAATTAACAGATCTTTAGAGCT 1080
 DB 1024 CAAGATCCCAAAAGTTCTGCTGAGATTTGATGGAATTAACAGATCTTTAGAGCT 1083
 QY 1081 CAATACTCGGCTGATCAAGTGAATGATGATGATGATGATGATGATGATGATGATGAT 1140
 DB 1084 CAATACTCGGCTGATCAAGTGAATGATGATGATGATGATGATGATGATGATGATGAT 1143
 QY 1141 AGCTTTTATCACTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1200
 DB 1144 AGCTTTTATCACTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1203
 QY 1201 TATGTTGGAGTTCCTCATTTTGGTATGATGATGATGATGATGATGATGATGATGATGAT 1260
 DB 1204 TATGTTGGAGTTCCTCATTTTGGTATGATGATGATGATGATGATGATGATGATGATGAT 1263

QY 1261 AGGAGAGCTGTGATGAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1320
 DB 1264 GGGAGAGAGCTGTGATGAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1323
 QY 1321 GAGAGAGCTGTGATGAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1380
 DB 1324 GAGAGAGCTGTGATGAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1383
 QY 1381 CCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1440
 DB 1384 CCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1443
 QY 1441 CCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1500
 DB 1444 CCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1503
 QY 1501 CTCTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1560
 DB 1504 CTCTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1563
 QY 1561 AAAATGTTTATTTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1620
 DB 1564 AAAATGTTTATTTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1620
 QY 1621 ATAGATCTTTC 1631
 DB 1621 TTAATTAAGTTC 1631

RESULT 6
HSUPGT
DEFINITION Human mRNA for liver microsomal UDP-glucuronosyltransferase (UDPGT).
ACCESSION Y00317
VERSION Y00317.1 GI:37588
KEYWORDS transferase; UDP-glucuronosyltransferase.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE 1
AUTHORS Jackson, M.R., McCarthy, L.R., Harding, D., Wilson, S., Coughtrie, M.W. and Burchell, B.
TITLE Cloning of a human liver microsomal UDP-glucuronosyltransferase cDNA
JOURNAL Biochem. J. 242 (2), 581-588 (1987)
MEDLINE 87241362
PUBMED 3109396
REFERENCE 2 (bases 1 to 2093)
AUTHORS Burchell, B.
TITLE Direct Submission
JOURNAL Submitted (26-JAN-1987) Burchell B., Dept. Biochemistry, Medical Sciences Institute, University Dundee, Scotland
FEATURES location/Qualifiers
source 1..2093
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="HUG25"
 /tissue_type="liver"
 /clone_lib="lambda gcl1 liver cDNA"
 38..1624
 /note="precursor"
 /codon_start=1
 /protein_id="CA68415.1"
 /db_xref="GI:37588"
 /db_xref="SWISS-PROT:P06133"
 /translation="MSMKWTSALLIOLSCVPSGSGCKVLVMPTEESHMNTKITLD
 ELVORGEVTVLASASISIFDPSPTLKEVTPVSLTKTEFDIITKOLVKRAELPK
 DTFWSYFSVOELIMTENDILRFKCDIVSNKKLMDKESRDVLAADVPPGELL
 AELIKLPVVRPSPGVAIEKSGGLPEPSPVPMVMSLSDQMTFIEIKVMIVL
 YFEFWFIQIDMKKMDQFYSFVLGRPTTLSETMAKADIWILIRNWDFOFPHPLPNEF

VGLHCKPAKPLPKMEEFVOSGNGVVFSLGSMVSNITSEBRANVASLAKIPK
 VLMPEDKPTLGLANTLKYKIPONDILKPTKRAFTTHGANGIYKAIPIPMVG
 VLPADOPDNIAMHKAKGAAYSLDEHTMSSDILNALKTIVINDLKYENAKLSRIH
 DQVYKPDRAVFIPIEFVPRKHAKHLRAADLTWFOHSLDVTGFLACVATVIFII
 TKCLFCVWKVFRTEKGRD"
 38.106
 /note="put. signal peptide"
 107.1621
 /product="mature UDPgt (AA 1-505) (EC 2.4.1.17)"
 979.998
 /note="pot. glycosylation site"
 2076.2081
 /note="pot. polyA signal"
 BASE COUNT 638 a 397 c 439 g 619 t
 ORIGIN
 Query Match 86.8%; Score 1432.6; DB 9; Length 2093;
 Best Local Similarity 92.8%; Pred. No. 0;
 Matches 1514; Conservative 0; Mismatches 114; Indels 3; Gaps 1;

1 AGCACTGAAAACAGATTCATGATGAGATGCTAATGAATGAGCTTCAGCTCT 60
 4 AGCACTGAAAACAGATTCATGATGAGATGCTAATGAATGAGCTTCAGCTCT 63
 61 TCTCTGATACAGCTGATGCTGTTAGCTCTGAGAGTTGTGGAAGAGTCTGTGTG 120
 64 TCTGCTGATACAGCTGATGCTGTTAGCTCTGAGAGTTGTGGAAGAGTCTGTGTG 123
 121 GCCCAGAAATTCAGCCACTGAGTAATTAAGCAATCTGTGATGAATCTGTCCAGAG 180
 124 GCCCAGAAATTCAGCCACTGAGTAATTAAGCAATCTGTGATGAATCTGTCCAGAG 183
 181 AGGTATAGAGGATGATGATGAGTATGAGTATGAGTATGAGTATGAGTATGAGTAT 240
 184 AGGTATAGAGGATGATGATGAGTATGAGTATGAGTATGAGTATGAGTATGAGTAT 243
 241 ATCTACTCTTAAATTTGAAGTTATCTGTATCTTTAACTTAACTGATTTGAGATAT 300
 244 ATCTACTCTTAAATTTGAAGTTATCTGTATCTTTAACTTAACTGATTTGAGATAT 303
 301 TATCAAGAGCTGTGTTAAGAGTGGCAGAACTTCCAAAAGACATTTTGTCTATATT 360
 304 TATCAAGAGCTGTGTTAAGAGTGGCAGAACTTCCAAAAGACATTTTGTCTATATT 363
 361 TTTCAAGAGCTGTGTTAAGAGTGGCAGAACTTCCAAAAGACATTTTGTCTATATT 420
 364 TTTCAAGAGCTGTGTTAAGAGTGGCAGAACTTCCAAAAGACATTTTGTCTATATT 423
 421 TATAGTTCAATAAGAACTTATAGAGAACTACAGAGAGTCAAGATTGATGTTCT 480
 424 TATAGTTCAATAAGAACTTATAGAGAACTACAGAGAGTCAAGATTGATGTTCT 483
 481 TGCAGATGCTGTTTCCCTTTGTGAGCTGTGAGCTGCGCAGATTCTTAAATACCTTTGT 540
 484 TGCAGATGCTGTTTCCCTTTGTGAGCTGTGAGCTGCGCAGATTCTTAAATACCTTTGT 543
 541 CTACAGCTCCGCTCTCTCCCTGCTAGCAGATTTAAAGACATTTAGAGAGCTCTGT 600
 544 CTACAGCTCCGCTCTCTCCCTGCTAGCAGATTTAAAGACATTTAGAGAGCTCTGT 603
 601 CCTCTCTCTATGAGCTGTTGTTATGTCAGAACTAAGTGAACCAATGACTTCATAGA 660
 604 CCTCTCTCTATGAGCTGTTGTTATGTCAGAACTAAGTGAACCAATGACTTCATAGA 663
 661 GAGGTAATAAATATGATCTATGCTTATTTGAATTTGGTTCCAAATATTTGACAT 720
 664 GAGGTAATAAATATGATCTATGCTTATTTGAATTTGGTTCCAAATATTTGACAT 723
 721 GAGGAAGTGGATGATGTTCTACAGGAAGTTCTAGGAAGACCCACTAGCTTATCTGAAG 780
 724 GAGGAAGTGGATGATGTTCTACAGGAAGTTCTAGGAAGACCCACTAGCTTATCTGAAG 783
 781 AATGCAAAAGCTGATATGCTTATTTGAAGAACTACTGAGATTTCAATTTCTCACCC 840

784 AATGCAAAAGCTGATATGCTTATTTGAACTACTGAGATTTCAATTTCTCACCC 843
 841 ACTTTACCAAAATTTGATGTTCTGTTGAGAGCTCCACTGCAAAACCTGCCAAACCCCTACC 900
 844 ACTTTACCAAAATTTGATGTTCTGTTGAGAGCTCCACTGCAAAACCTGCCAAACCCCTACC 903
 901 GAGGAAGTGGATGATGTTCTACAGGAAGTTCTAGGAAGACCCACTAGCTTATCTGAAG 960
 904 GAGGAAGTGGATGATGTTCTACAGGAAGTTCTAGGAAGACCCACTAGCTTATCTGAAG 963
 961 GGGGTGATGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1020
 964 GGGGTGATGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1023
 1021 CAAGATCCCAAAAGCTGTTGTTGAGATTTGATGAGAAATTAACAGATTTAGAGACT 1080
 1024 CAAGATCCCAAAAGCTGTTGTTGAGATTTGATGAGAAATTAACAGATTTAGAGACT 1083
 1081 CAATATCTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1140
 1084 CAATATCTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1143
 1141 AGCTTTATATCACTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1200
 1144 AGCTTTATATCACTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1203
 1201 TATGTTGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1260
 1204 TATGTTGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1263
 1261 AGGAGCAGCTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1320
 1264 AGGAGCAGCTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1323
 1321 GAGAAAGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1380
 1324 GAGAAAGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1383
 1381 CCATGAT 1440
 1384 CCATGAT 1443
 1441 CCATGAT 1500
 1444 CCATGAT 1503
 1501 CTCTATGAT 1560
 1504 CTCTATGAT 1563
 1561 AAAATGTTTATTTTCTGCTGCAAAAATTTAATTAAGTGAAGAAAGAGAGGA 1620
 1564 AAAATGTTTATTTTCTGCTGCAAAAATTTAATTAAGTGAAGAAAGAGAGGA 1620
 1621 AATGATCTTTC 1631
 1621 TTAATTAAGT 1631

RESULT 7
 AF081793
 LOCUS
 DEFINITION Homo sapiens UDP glucuronosyltransferase 284 precursor (UGT2B4)
 mRNA, UGT2B4*V109, U396 variant allele, complete cds.
 ACCESSION AF081793
 VERSION AF081793.1 GI:3426331
 KEYWORDS
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 2091)

AUTHORS Jin, C.J., Miners, J.O., Lillywhite, K.J. and Mackenzie, P.I.
TITLE cDNA cloning and expression of two new members of the human liver
UDP-glucuronosyltransferase 2B subfamily
JOURNAL Biochem. Biophys. Res. Commun. 194 (1), 496-503 (1993)
MEDLINE 83326164
PubMed 8333863
REFERENCE 2 (bases 1 to 2091)
AUTHORS Mackenzie, P.I.
TITLE Direct Submision
JOURNAL Submitted (30-JUL-1998) Clinical Pharmacology, Flinders University
of South Australia, Bedford Park, SA 5042, Australia
FEATURES
source
1. .2091
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/chromosome="4"
1. .2091
/gene="UGT2B4"
/note="formerly UGT2B11"
/allele="UGT2B4*1109, L396 variant"
36. .1622
/gene="UGT2B4"
/codon_start=1
/product="UDP glucuronosyltransferase 2B4 precursor"
/protein_id="AAC32272.1"
/db_xref="GI:3426332"
/translation="MSMKRTSALLLIQLSCYFSSGSGKVLVMPFESHMNTITLD
ELVORGHVTVLASSASTISDPNSPSTLKEFVPSLKTKEFDITLQVLRWMLPK
DTEWYSLSQVOELIMTENDILRKFCDIVSKMLKQESRFPVLAADAFPEGLL
AELKILFPIVSLRSPGVALEKSHGSLFPPSPVPMVSELSDDMTFIERKMIYVL
YFEFROIPEMKKMDQFSEVLGRPTLSEFMKADLIJENRYADFPQPHLPNVEF
VGGHCHCPARPLPKMKKEFVSGENGVVSLGSMVSNSEERANYIASLAKTPRK
VLMFEDNKPTPLGLNRLYKMTIPQNDLGPRTKRAFTTHGANGIYEAIRHGPWG
VPLIADQDNTAHMKAKGAASVLDFTWSTDLNLAKTVINDPLVRENAKLSRIH
DQPKPLDRAVFWIEFWRHKGAHLRVAADHLMFQYHSLDVTGFLACVATVFIIT
TKCLFCWKFRTKGKGRD"
BASE COUNT 638 a 401 c 437 g 615 t
ORIGIN
Query Match 86.8%; Score 1431.6; DB 9; Length 2091;
Best Local Similarity 92.8%; Pred. No. 0;
Matches 1513; Conservative 0; Mismatches 114; Indels 3; Gaps 1;
QY 2 GCAACTGGAACAACAAGCATTCATTCAGATGCTGTGAATGAGACTTCAGCTTT 61
DB 3 GCAACTGGAACAACAAGCATTCATTCAGATGCTGTGAATGAGACTTCAGCTTT 62
QY 62 CTCTGATACAGCTGAGCTGTTACTTGTAGCTCTGGAGATTGTGAAAGTCTGGTGG 121
DB 63 CTGCTGATACAGCTGAGCTGTTACTTGTAGCTCTGGAGATTGTGAAAGTCTGGTGG 122
QY 122 CCCACAGATTACAGCCATGATGATTAATAAGCAATCTGTGAACTTGTCCAGGA 181
DB 123 CCCACAGATTACAGCCATGATGATTAATAAGCAATCTGTGAACTTGTCCAGGA 182
QY 182 GGTATAGAGTGATCTGATTTGGCATCTTCAGCTTCATTTCTTTCGATCCCAAGCCCA 241
DB 183 GGTATAGAGTGATCTGATTTGGCATCTTCAGCTTCATTTCTTTCGATCCCAAGCCCA 242
QY 242 TCTACTCTTAATTTGAAGTTATCTGTATCTTAACTTAAGAGTGAAGATTT 301
DB 243 TCTACTCTTAATTTGAAGTTATCTGTATCTTAACTTAAGAGTGAAGATTT 302
QY 302 ATCAAGCAGCTGTTAAGAGATGGCAGAACTTCCAAAAGACACATTTTGTCAATTTT 361
DB 303 ATCAAGCAGCTGTTAAGAGATGGCAGAACTTCCAAAAGACACATTTTGTCAATTTT 362
QY 362 TCACAAGTACAAAGAAATATGTGACATTTAATGACATCTTAAGAAAGTTCTGTAAGAT 421
DB 363 TCACAAGTACAAAGAAATATGTGACATTTAATGACATCTTAAGAAAGTTCTGTAAGAT 422
QY 422 ATAGTTTCAATTAAGAACTTATGAAAGAACTACAGAGTCAAGATTTGATGTTCTT 481

DB 423 ATAGTTTCAATTAAGAACTTATGAAAGAACTACAGAGTCAAGATTTGATGTTCTT 482
QY 482 GCAATCTGCTTTTCCCTTCTTGGAGACTGCTGGCCAGTACTTAATAATCCCTTTGTC 541
DB 483 GCAAGCCTGCTTTTCCCTTCTTGGAGACTGCTGGCCAGTACTTAATAATCCCTTTGTC 542
QY 542 TACAGCTCGCTTCTCTCTGAGTACGCAATTAAGAGCATAGTGAAGACCTCTGTC 601
DB 543 TACAGCTCGCTTCTCTCTGAGTACGCAATTAAGAGCATAGTGAAGACCTCTGTC 602
QY 602 CTCTCTTCTATGTCCTGTTGTTTATGTCAGAACTAAGTGAACCAATGACTTTCATAGAG 661
DB 603 CTCTCTTCTATGTCCTGTTGTTTATGTCAGAACTAAGTGAACCAATGACTTTCATAGAG 662
QY 662 AGGTTAAAAAATATGATCTATGCTTATTTGAAATTTTGTCCAAATTTTGAACATG 721
DB 663 AGGTTAAAAAATATGATCTATGCTTATTTGAAATTTTGTCCAAATTTTGAACATG 722
QY 722 AAGAAGTGGGATCAGTCTACAGTGAAGTCTAGGAAGACCCACTAGCTTATCTGAGACA 781
DB 723 AAGAAGTGGGATCAGTCTACAGTGAAGTCTAGGAAGACCCACTAGCTTATCTGAGACA 782
QY 782 ATGGCAAAAGCTGACATATGAGCTTATTCGAAACTACTGGAATTTTCAATTTCTCACCCA 841
DB 783 ATGGCAAAAGCTGACATATGAGCTTATTCGAAACTACTGGAATTTTCAATTTCTCACCCA 842
QY 842 CTCTTACCAATGTTAGTCTGTTGGAGACTCCACTGCAAACTGCGCAAACTCTACCG 901
DB 843 CTCTTACCAATGTTAGTCTGTTGGAGACTCCACTGCAAACTGCGCAAACTCTACCG 902
QY 902 AAGGAATGGAAGATTTGTCAGAGCTCGAGAAATAGTGTGTGTTGTTTCTCTG 961
DB 903 AAGGAATGGAAGATTTGTCAGAGCTCGAGAAATAGTGTGTGTTGTTTCTCTG 962
QY 962 GGGTGCATGTCAGTAAACAGTCAAGAAAGGCGCAATGATTAATGATCAGCCCTTGCC 1021
DB 963 GGGTGCATGTCAGTAAACAGTCAAGAAAGGCGCAATGATTAATGATCAGCCCTTGCC 1022
QY 1022 AAGATCCCAAAAGGTTCTGTGAGATTTGATGGAAATTAACCAATCTTTAGACTC 1081
DB 1023 AAGATCCCAAAAGGTTCTGTGAGATTTGATGGAAATTAACCAATCTTTAGACTC 1082
QY 1082 AATACCTGGCTGTAACAAGTGAATCCCAAGATATCTTGTGTCACCCAAAACCAA 1141
DB 1083 AATACCTGGCTGTAACAAGTGAATCCCAAGATATCTTGTGTCACCCAAAACCAA 1142
QY 1142 GCTTTATCACTCATGTGTGAATGAATGGAATCTATGAAGCTATTTCAATGGGTCCT 1201
DB 1143 GCTTTATCACTCATGTGTGAATGAATGGAATCTATGAAGCTATTTCAATGGGTCCT 1202
QY 1202 ATGGTGGAGTTCCTCATATTTTGTGATGACGCTGATTAACATAGTCAATGAGCCAAA 1261
DB 1203 ATGGTGGAGTTCCTCATATTTTGTGATGACGCTGATTAACATAGTCAATGAGCCAAA 1262
QY 1262 GAGACAGCTGTAGAAATTAACCTCAAACTATGACAAGGCAAGTTACTGAGGGCTTTG 1321
DB 1263 GAGACAGCTGTAGAAATTAACCTCAAACTATGACAAGGCAAGTTACTGAGGGCTTTG 1322
QY 1322 AGAAGATCTTACCGATTCCTCTTATTAAGAGATCTATGAGATTAACAAGATTTAC 1381
DB 1323 AAGACATATTAATGATTCCTTTATTAAGAGATCTATGAGATTAACAAGATTTAC 1382
QY 1382 CATGATCACTGTAAAGCCCTAGATGAGACGCTTCTGATGAGCTTTGATGCGC 1441
DB 1383 CATGATCACTGTAAAGCCCTAGATGAGACGCTTCTGATGAGCTTTGATGCGC 1442
QY 1442 CACAAAGAGCCAAAGCCTGCGATGAGCTGCCATGACCTGCTGCTGAGCACTAC 1501
DB 1443 CATTAAGAGCCAAAGCCTTGGGTTGACGCGCCACGACCTGACCTGCTGCTGAGCACT 1502
QY 1502 TCTATGATGATTTGGGTTCTGCTGACCTGCTGCTGAGCACTGCTATATTTCTGTTACA 1561
DB 1502 TCTATGATGATTTGGGTTCTGCTGACCTGCTGCTGAGCACTGCTATATTTCTGTTACA 1561

Db 1503 TCCTTGATGACTGGCTCCTGCGCTGCTGTCGCACTGATATTCATCATCA 1562
Qy 1562 AAAGCTTTTATTTTCTGTCAAAATTAATAAGATGAAAGGGAA 1621
Db 1563 AAATGCTGTTTGTGCTG--GAAGTTTGAACAAGAAAGAGAGAT 1619
Qy 1622 TAGATCTTTC 1631
Db 1620 TAATTACCTC 1629

RESULT 8
AX548042 1639 bp DNA Linear PAT 26-NOV-2002
LOCUS Sequence 18 from Patent WO0206654.
DEFINITION AX548042
ACCESSION AX548042.1 GI:25813138
VERSION
KEYWORDS
SOURCE
ORGANISM Homo sapiens (human)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
REFERENCE
AUTHORS 1
Astromoff, A., Au-Young, J., Baughn, M.R., Ding, L., Duggan, B.M.,
Forsey, I.J., Gietzen, K.J., Griffin, J.A., Lee, E.A., Lu, Y.,
Richardson, T.W., Ring, H.Z., Sanjanwala, M.M., Swarnakar, A.,
Walia, N.K., Warren, B.A., Xu, Y., Yue, H. and Zebardjian, Y.
TITLE Drug metabolizing enzymes
JOURNAL Patent: WO 0206654-A 18 29-AUG-2002;
Incyte Genomics, Inc. (US)
FEATURES
SOURCE 1. 1639
Location/Qualifiers
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/note="Incyte ID No: 7493833CB1"

BASE COUNT 470 a 316 c 372 g 481 t
ORIGIN

Query Match 80.7%; Score 1331.4; DB 6; Length 1639;
Best Local Similarity 89.6%; Pred. No. 0;
Matches 1431; Conservative 0; Mismatches 166; Indels 0; Gaps 0;

Qy 35 ATGTATGAATGAGCTTCACTCTCTCTCTGATATACAGCTGAGCTGTTACTTACTCT 94
Db 1 ATGTATGAATGAGCTTCACTCTCTCTCTGATATACAGCTGAGCTGTTACTTACTCT 60
Qy 95 GGGAGTTGTGAAAGGTGCTGTGCGCCACAGAAATTCAGCACTGATGAATAAAG 154
Db 61 GGGAGTTGTGAAAGGTGCTGTGCGCCACAGAAATTCAGCACTGATGAATAAAG 120
Qy 155 ACAATCTCGATGAACCTGTCCAGAGAGTCAATGAGTGAATGAGTCACTTCACT 214
Db 121 ACAATCTCGATGAACCTGTCCAGAGAGTCAATGAGTGAATGAGTCACTTCACT 180
Qy 215 TCCATTTCTTGATCCCAAGCCCATCTACTCTTAATTTGAAGTTATCTGATCT 274
Db 181 TCCATTTCTTGATCCCAAGCCCATCTACTCTTAATTTGAAGTTATCTGATCT 240
Qy 275 TTAATAAAGCTGATTTGAGATATTAACAAGAGCTGTGTTAAGATGGCAGAACTT 334
Db 241 TTAATAAAGCTGATTTGAGATATTAACAAGAGCTGTGTTAAGATGGCAGAACTT 300
Qy 335 CCAAAAGACACTTTTGTGATATTTTACAAGTAACAAGAAATCATGTGACATTTAT 394
Db 301 CCAAAAGACACTTTTGTGATATTTTACAAGTAACAAGAAATCATGTGACATTTAT 360
Qy 395 GACATACCTAGAAAGTTCTGTAAGTATAGTTTCAATTAAGAACTTATGAAGAACTA 454
Db 361 GACATACCTAGAAAGTTCTGTAAGTATAGTTTCAATTAAGAACTTATGAAGAACTA 420
Qy 455 CAGAGTCAAGATTTGATGTTCTTTCGAGATGCTTTCCCTTTGAGAGTCTG 514

Db 421 CAGAGTCAAGATTTGACGTCATTTTTCAGAGTCAATTTTCCCTGTAGTACTGCTG 480
Qy 515 GCGAGTACTTAATAATACCTTTGTGTCAAGCTCGCTCTCTCTGAGTCAAGAT 574
Db 481 GCTAGCTATTTTAACATACCTTTGTGTCAAGCTCGCTCTCTCTGAGTCAAGAT 540
Qy 575 GAAAGCATAGTGAAGACTTCTGCTCCCTCTCTCTATGCTGCTGTTTATGCAAA 634
Db 541 GAAAGCATAGTGAAGACTTATTTTCCCTCTCTCTATGCTGCTGTTTATGCAAA 600
Qy 635 CTAAGTCAACCAATGACTTTTATAGAGAGGTAAATAATGATCTATGCTTATTT 694
Db 601 TTAATGATCAAAAGACTTTTATGAGAGAGGTAAATAATGATCTATGCTTATTT 660
Qy 695 GAATTTGTTCCAAATATTTGACATGAAGAGTGGATCAGTTCTACAGTAAGTTCTA 754
Db 661 GACTTTGTTCCAAATATTTGACATGAAGAGTGGATCAGTTCTATAGTAAGTTCTA 720
Qy 755 GAAAGCCCACTAGCTTATCTAGACAATGGAAGCTGACATATGCTTATTCGAAAC 814
Db 721 GAAAGCCCACTAGCTTATCTAGACAATGGAAGCTGACATATGCTTATTCGAAAC 780
Qy 815 TACTGGATTTTCAATTTCTCAACCTCTTACCAATGTTGAGTCTGTTGAGACTC 874
Db 781 TCTGGAATTTTCAATTTCTCAACCTCTTACCAATGTTGAGTCTGTTGAGACTC 840
Qy 875 CACTGCAAACTGCGCAAAACCCCTACCGAAGGAATGGAAGTTGTCAGAGCTTGA 934
Db 841 CACTGCAAACTGCGCAAAACCCCTACCGAAGGAATGGAAGTTGTCAGAGCTTGA 900
Qy 935 GAAATGTTGTTGTTGTTTCTCTGAGGATGATGCTGATACAGCTGCAAGAAAG 994
Db 901 GAAATGTTGTTGTTGTTTCTCTGAGGATGATGCTGATACAGCTGCAAGAAAG 960
Qy 995 GCCAATGATTTGATCAGCCCTTCCCAAGATCCCAAAAGTTCTCTGAGAAATTTGAT 1054
Db 961 GCCAATGATTTGATCAGCCCTTCCCAAGATCCCAAAAGTTCTCTGAGAAATTTGAT 1020
Qy 1055 GGAATTAACAGATATCTTGAAGTCAATCTCGGCTGATGATGATGATGATGAT 1114
Db 1021 GGAATTAACAGATATCTTGAAGTCAATCTCGGCTGATGATGATGATGATGAT 1080
Qy 1115 GATCTTCTTGTGATCCCAAAACCAAGCTTTTATCACTCATGTTGATGAATGATGAT 1174
Db 1081 GATCTTCTTGTGATCCCAAAACCAAGCTTTTATCACTCATGTTGATGAATGATGAT 1140
Qy 1175 TATGAAGTATTTACATGAGGCTCTATGTTGAGAGTTCCCATTTTGGATGATGAT 1234
Db 1141 TATGAAGTATTTACATGAGGCTCTATGTTGAGAGTTCCCATTTTGGATGATGAT 1200
Qy 1235 GATAACATAGCTCATGAAGGCGCAAGAGAGAGCTGATGAATTAACCTTCAAACTATG 1294
Db 1201 GATAACATAGCTCATGAAGGCGCAAGAGAGAGCTGATGAATTAACCTTCAAACTATG 1260
Qy 1295 ACAAGCGAAGATTTACTGAGGCTTTGAGAA CAGCTATTAACGATTCCTTTAAGAG 1354
Db 1261 TCGAGTACAGCTGCTGAATGAATGAAGAGATTAATTAATCTTCAATATGAAGAG 1320
Qy 1355 AATGCTATGAATTTACAAAGATTCACATGATCAACCTGTAAGGCCCCCTGATGACCA 1414
Db 1321 AATGCTATGAATTTACAAAGATTTCAACATGATCAACAGTAAGGCCCCCTGATGACCA 1380
Qy 1415 GTCTTCTGATGAGTTTGTGATGAGGCGCAAGAGAGAGCAAGCACTGATGAGCTGCC 1474
Db 1381 GTCTTCTGATGAGTTTGTGATGAGGCGCAAGAGAGAGCAAGCACTGATGAGCTGCC 1440
Qy 1475 CATGACCTACCTGTTCCAGACACTCTATATGATGATGATGATGATGATGATGATGAT 1534
Db 1441 CACGACCTACCTGTTCCAGACACTCTTATGATGATGATGATGATGATGATGATGATGAT 1500
Qy 1535 GTGCACTGCTATATTTCTTGTCAAAAGTTTATTTTCTGTCGCAAAATTTAT 1594
Db 1501 GTGCACTGCTATATTTCTTGTCAAAAGTTTCTGTTTGTGTTGTTGTTGTTGTTGTT 1560

Qy 1595 AAAAGTAAAGATGAGAAAGAGATGATCTTTC 1631
Db 1561 AGAAGAGAAAGAGGAGAAAGATTAATTACGTC 1597

RESULT 9
AF112112 2108 bp mRNA linear PRI 11-APR-1999
LOCUS Macaca fascicularis UDP-glucuronosyltransferase 2B19 precursor,
AF112112
DEFINITION mRNA, complete cds.
ACCESSION AF112112
VERSION AF112112.1 GI:4580601
KEYWORDS Macaca fascicularis (crab-eating macaque)
SOURCE Macaca fascicularis
ORGANISM Macaca fascicularis
Bukariota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
Cercopithecinae; Macaca.
REFERENCE 1 (bases 1 to 2108)
AUTHORS Belanger, G., Barbier, O., Hum, D.W. and Belanger, A.
TITLE Molecular cloning, expression and characterization of a monkey
steroid UDP-glucuronosyltransferase, UGT2B19, that conjugates
testosterone
JOURNAL Eur. J. Biochem. 260 (3), 701-708 (1999)
MEDLINE 99203465
PUBMED 10102998
REFERENCE 2 (bases 1 to 2108)
AUTHORS Belanger, G., Barbier, O., Hum, D.W. and Belanger, A.
TITLE Direct Submission
JOURNAL Submitted (07-DEC-1998) Molecular Endocrinology, CHUL Research
Center, 2705 Laurier Blvd, Sainte Foy, Quebec G1V 4G2, Canada

FEATURES
source location/Qualifiers
1..2108
/organism="Macaca fascicularis"
/mol_type="mRNA"
/db_xref="taxon:9541"
26..1612
/EC_number="2.4.1.17"
/note="UGT2B19"
/codon_start=1
/product="UDP-glucuronosyltransferase 2B19 precursor"
/protein_id="AAD2435.1"
/db_xref="GI:4580602"
/translation="MSMKWTSAALLLIOLSCYLSFGSGKULWPFESHMNNITIID
ELVORGHEVTVLAYSTSLDPNNPSPKFEICPTSLTEFQDSVQLVKRMSDIK
DTEWPHFLHVEEMWYTGDMIRKFCQOVNSKMLKLOESRFVVLADALSISGELL
AELIKIPFVYSILRSPGVALEKHGGGLFPPSYVPVIMSELDMERVONMIYVY
YFDFWFOVMDVKNDOFASKVLGRPTLFEIMAKAELIMLIYNDFOEPHLLPNEVE
VCGIHCAPKRLPKREMEFVSGDNCVVFSLGSMVSNSEEPANVIAASLAKIPK
VLWRFDPNKPTLGLNTOLYKMLPONDILGHPKTRATITGGAGATYBAIYHGIPIWG
VPLFADQPDNTAHKAKGAARLDPTMSSLDLINALKTINDIYKENARKLSIHH
DQVPLDRAVFWIEFWRRHKGAGHLVAAHDLTFWHSIDVIGPLACVATVIFII
TKCLFCVWKFEVTRKKGRD"

BASE COUNT 649 a 393 c 438 g 628 t
ORIGIN

Query Match 78.3%; Score 1292.4; DB 9; Length 2108;
Best Local Similarity 87.7%; Pred. No. 0;
Matches 1423; Conservative 0; Mismatches 196; Indels 3; Gaps 1;

Qy 10 AAAACAGCATTGATTCATCAGATGTCATGAAATGACCTTCAGCTCTTCCCTGAT 69
Db 1 AGAACAAGCATTGATTCATCAGATGTCATGAAATGACCTTCAGCTCTTCCCTGAT 60

Qy 70 ACACCTGAGCTGTACTTACTGAGCTGGAGTGTGAAAGTGCTGTGTGGCCACAGA 129
Db 61 ACACCTGAGCTGTACTTACTGAGCTGGAGTGTGAAAGTGCTGTGTGGCCACAGA 120

Qy 130 ATTGAGCCTGATGATTAATAAGACAATCTTGATGAATCTTGTCCAGAGAGGTATGA 189
Db 121 ATTGAGCCTGATGATTAATAAGACAATCTTGATGAATCTTGTCCAGAGAGGTATGA 180

Qy 190 GGTGACTGATATGGCATCTTCAGCTTCATTTCTTTGGATCCCAAGACCCTACTCT 249
Db 181 GGTGACTGATATGGCATCTTCAGCTTCATTTCTTTGGATCCCAAGACCCTACTCT 240

Qy 250 TAAATTTGAAGTTATCTGATCTTTTAACTAACTGAGTTGAGATATTAACAGCA 309
Db 241 TAAATTTGAAGTTATCTGATCTTTTAACTAACTGAGTTGAGATATTAACAGCA 300

Qy 310 GGTGGTTAAGATGGGCGAGAACTTCCAAAGACATTTTGGTCATATTTTTCACAGT 369
Db 301 ACTGGTTAAGATGGGCGAGAACTTCCAAAGACATTTTGGGCGACATTTTTCACAGT 360

Qy 370 ACAAGAAATCATGTGACATTTAATGACATATGAAAGTTCTGTAAGATATAGTTTC 429
Db 361 ACAAGAAATCATGTGACATTTAATGACATATGAAAGTTCTGTAAGATATAGTTTC 420

Qy 430 AAATAGAACTTATGAGAAACTACAGAGTCAAGATTGATGTTCTTGCAGATGC 489
Db 421 TAAATAGAACTTATGAGAAACTACAGAGTCAAGATTGATGTTCTTGCAGATGC 480

Qy 490 TGTTCCTCCCTTGGTGGAGCTGCGCGGAGTTAATAAATCCCTTGTCTACAGCT 549
Db 481 TATTCCTCCCTGTTGGTGGAGCTGCGCGGAGTTAATAAATCCCTTGTCTACAGCT 540

Qy 550 CCGCTTCTCTCGGCTACGCAATTGAAAGACATAGTGAAGACTTCTGTTCCCTCTTC 609
Db 541 CCGCTTCTCTCGGCTACGCAATTGAAAGACATAGTGAAGACTTCTGTTCCCTCTTC 600

Qy 610 CTATGTCCTGTTGTTATGTCAGAACTAAGTCAAAATGATCTTCTATGAGAGGTAA 669
Db 601 CTATGTCCTGTTGTTATGTCAGAACTAAGTCAAAATGATCTTCTATGAGAGGTAA 660

Qy 670 AAATATGATCTATGCTTTTATTTTGAATTTGGTTCCAAATATTTACATGAAGATG 729
Db 661 AAATATGATCTATGCTTTTATTTTGAATTTGGTTCCAAATATTTACATGAAGATG 720

Qy 730 GATATGATCTATGCTTTTATTTTGAATTTGGTTCCAAATATTTACATGAAGATG 789
Db 721 GATATGATCTATGCTTTTATTTTGAATTTGGTTCCAAATATTTACATGAAGATG 780

Qy 790 AGCTGACATATGCTTTATTTGAAACTAAGTGGATTTTCAATTTCTGACCCACTTTAC 849
Db 781 AGCTGACATATGCTTTATTTGAAACTAAGTGGATTTTCAATTTCTGACCCACTTTAC 840

Qy 850 AATATGATGCTGTTGAGAGACTCCACTGCAAACTGCGCAAACTCCAGAGAAAT 909
Db 841 AATATGATGCTGTTGAGAGACTCCACTGCAAACTGCGCAAACTCCAGAGAAAT 900

Qy 910 GGAAGATTTGTCAGAGCTCTGAGAAATAGGTGTGTGTTTCTCTGGGGTGCAT 969
Db 901 GGAAGATTTGTCAGAGCTCTGAGAAATAGGTGTGTGTTTCTCTGGGGTGCAT 960

Qy 970 GGTGATGACAGCTCAGAGAAAGGCGCAATGTATTCATCAGCCCTTGCAAGATCCC 1029
Db 961 GGTGATGACAGCTCAGAGAAAGGCGCAATGTATTCATCAGCCCTTGCAAGATCCC 1020

Qy 1030 ACAAAAGCTTCTGTGAGATTTGATGGAGATTAACCAAGATCTTATAGATCTCAATCTCG 1089
Db 1021 ACAAAAGCTTCTGTGAGATTTGATGGAGATTAACCAAGATCTTATAGATCTCAATCTCA 1080

Qy 1090 GCTGTACAGATGAGTACCCAGATGATCTTCTTGGTCATCCCAAAACCAAGCTTTTAT 1149
Db 1081 GCTGTACAGATGAGTACCCAGATGATCTTCTTGGTCATCCCAAAACCAAGCTTTTAT 1140

Qy 1150 CACTCATGTGATGATGATGAGATCTATGAAGTATTTACATGAGGCTCCCTATGATGG 1209
Db 1141 AACTCATGTGATGATGAGATCTATGAAGTATTTACATGAGGCTCCCTATGATGG 1200

Qy 1210 AGTTCCCATATTTGGTATGATGATGATGATGATGATGATGATGATGATGATGATG 1269
Db 1201 CGTTCCCATATTTGGTATGATGATGATGATGATGATGATGATGATGATGATGATG 1260

Qy 1270 TGTAGAAATTAATCTTCAAACTATGACAGCAAGATTTACTGAGGCTTTGAGAACAT 1329

Db 1261 TGTAGACTGACCTTCGACACAAATGTGAGTACTTCAATGCACTGAAAGACAGT 1320
Qy 1330 CATTAACCATTCCTCTTTAAAGAGATGCTATGAGATTATCAAGATTCAACCATGATCA 1389
Db 1321 AATAAATGATCTTATATATTAAGAAATGCTAATAATTAATCAAGATTCAATGATCA 1380
Qy 1390 ACCTGTAAAGCCCCAGATGAGAGAGCTTCTGATGAGTTTGTATGATGCGCAAAAG 1449
Db 1381 ACCAGTGAAGCCCTCGATTCAGAGCTCTTCTGATTAATTTGTCATGCGCCATTAAG 1440
Qy 1450 AGCCAGACCTGCGATGAGTCCATGACCTCAGCTGTTCCAGACTATCTATAGA 1509
Db 1441 AGCCAAACACCTGGGGTTCAGCCATGACCTCAGCTGTTCCAGACTCTTTGGA 1500
Qy 1510 TGTGATTTGGTTCCTGCTGAGCTGTGTGGCAATGCTATTTCTTTTCAAAAATGTTT 1569
Db 1501 TGTGATTTGGTTCCTGCTGAGCTGTGTGGCAATGCTATTTCTTTTCAAAAATGTTT 1560
Qy 1570 TTTATTTTCTGTCAAAAATTTAATAAAGTAAAGATAGAAAGAGGATATGATCTT 1629
Db 1561 GTTTTGTCTG--GAGTTTGTGAACAAGAAAGAGGAAAGATTAATCA 1617
Qy 1630 TC 1631
Db 1618 TC 1619

RESULT 10

BC030974 1766 bp mRNA linear PRI 13-JUN-2002
LOCUS Homo sapiens, UDP glycosyltransferase 2 family, polypeptide B7,
DEFINITION clone MGC:32620 IMAGE:4593426, mRNA, complete cds.

ACCESSION BC030974
VERSION BC030974.1 GI:21411301
KEYWORDS MGC.

SOURCE Homo sapiens (human)
ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
1 (base 1 to 1766)

REFERENCE 1 Strausberg, R.
TITLE Direct Submission
JOURNAL Submitted (03-JUN-2002) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA

REMARK NIH-MGC Project URL: <http://mgc.nci.nih.gov>
COMMENT Contact: MGC help desk
Email: cgapbs-remail.nih.gov
Tissue Procurement: CLONTECH
cDNA Library Preparation: CLONTECH Laboratories, Inc.
cDNA Library Arrived by: The I.M.A.G.E. Consortium (LIML)
DNA Sequencing by: Sequencing Group at the Stanford Human Genome
Center, Stanford University School of Medicine, Stanford, CA 94305
Web site: <http://www-shgc.stanford.edu>
Contact: (Dickson, Mark) mdickpax11.stanford.edu
Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers,
R. M.

Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LIML at: <http://image.liml.gov>
Series: IRPL Plate: 41 Row: e Column: 8
This clone was selected for full length sequencing because it
passed the following selection criteria: matched mRNA gi: 4507824.

FEATURES

source
1..1766
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="LOCUSID:7364"
/db_xref="taxon:9606"
/clone="MGC:32620 IMAGE:4593426"
/tissue_type="Kidney"

CDS

/clone_1lb="NIH_MGC_75"
/lab_host="DH10B"
/note="vector: pDNR-LIB"
47..1636
/codon_start=1
/product="UDP glycosyltransferase 2 family, polypeptide
B7"
/protein_id="AAH30974.1"
/db_xref="GI:21411302"
/translation="MSYKMTSVLLIQLSPFSGNGCKVLAWAESYSHMNITKIL
ELIQRHEVIVLSSASILPDRNNSSLSKTEIYPTSLTKTELNFIMOOIKRMSDLK
DTEFWLPSYQVEIWSITFGDITRKCKOVSNKFKMKVQSRPVIADAIKPSSELL
AELENIIPYVLSFSFGYTEKSGGGLPEPSVYPVMSLTDQMTMEVKNKIYV
YFDFWEIIPMKMKDQYSEVLGPTLSETMGKADVWLIKNSNMFQFPYPLPNVD
VGLGHCRAKPLPEKEMEDFVSSGENGVVFSLSMSNMTSEKAVYASLAQIPQK
VWRFQDNKPDITGLNRLTKMIKPNLDLGHPRATITGGAAGIYEALYHGIPOK
IPLPADQPDNIAMHKKAAGAVRVDNMSSTDLINAKRIVINDPSYKEMTKSLRIG
DQVKEPLDRAVFWIEFVWRHKGALHLEVAALDLTFQYHSLDIYIGFLVCVATYIFIV
TKCCLPFQWFKARAKKQND"

BASE COUNT 546 a 327 c 381 g 512 t
ORIGIN

Query Match 76.8%; Score 1267.8; DB 9; Length 1766;
Best Local Similarity 86.1%; Pred. No. 26-313;
Matches 1404; Conservative 0; Mismatches 227; Indels 0; Gaps 0;

Qy 1 AGCAACTGGAAAAAAGACATTCATGATGAGATGCTATGAAATGACCTCAGCTCT 60
Db 13 AGCAACTGGAAAAAAGACATTCATGATGAGATGCTATGAAATGACCTCAGCTCT 72
Qy 61 TCTCTGATACAGCTGAGCTGTTTACTTTAGCTCTGGAGTTGGAAAGTGTGTG 120
Db 73 TTTCTAATACAACTGAGCTTTTCTTTAGCTCTGGAGATTTGGAAGGCTGTGTG 132
Qy 121 GCCACAGAAATTCAGCCACTGATGATTAATAACAATCTCGATGAACCTTGCCAG 180
Db 133 GGCACAGAAATTCAGCCACTGATGATTAATAACAATCTCGATGAACCTTGCCAG 192
Qy 181 AGGTCATGAGGTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 240
Db 193 AGGTCATGAGGTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 252
Qy 241 ATCTACTCTTAAATTTGAAGTTTATCTGTATCTTTAACTTAACTGAAGTTGAGATAT 300
Db 253 ATCTACTCTTAAATTTGAAGTTTATCTGTATCTTTAACTTAACTGAAGTTGAGATAT 312
Qy 301 TATCAAGAGCTGTTAAGATGAGCAAACTTCCAAAAGACATTTTGTGATATTT 360
Db 313 CATCATGACACAGATTAAGAGATGATGATGATGATGATGATGATGATGATGAT 372
Qy 361 TTCAAGATGACAAAGATCATGATGATGATGATGATGATGATGATGATGATGATGAT 420
Db 373 TTCAAGATGACAAAGATCATGATGATGATGATGATGATGATGATGATGATGATGAT 432
Qy 421 TATAGTTTCAATTAAGAACTTATGAAGAACTAAGAGTCAAGATTTGATGTTTCT 480
Db 433 TGTAGTTTCAATTAAGAACTTATGAAGAACTAAGAGTCAAGATTTGATGTTTCT 492
Qy 481 TGCAGATGCTGTTTCCCTTTGAGCTGCTGCGCAGATTAATAAACCTTTGT 540
Db 493 TGCAGATGCTGTTTCCCTTTGAGCTGCTGCGCAGATTAATAAACCTTTGT 552
Qy 541 CTACAGCTGCTGCTTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 600
Db 553 GTACAGCTGCTGCTTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 612
Qy 601 CCTCTCTTCTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 660
Db 613 CCTCTCTTCTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 672
Qy 661 GAGGTAATAAATATGATCTATGCTGCTTATTTGAAATTTGGTTCCAAATATTTGACAT 720
Db 673 GAGGTAATAAATATGATCTATGCTGCTTATTTGAAATTTGGTTCCAAATATTTGACAT 732

```

QY 721 GAAGAAGTGGATGATCTTCTACAGTGAAGTCTAGGAAGACCCTACGTTATCTGAGAC 780
Db 733 GAAGAAGTGGATGATCTTCTACAGTGAAGTCTAGGAAGACCCTACGTTATCTGAGAC 792
QY 781 AATGGCAAAAGCTGACATATGCTTATTCGAACCTACTGGGATTTTCAATTTCTCACC 840
Db 793 AATGGGAAAAGCTGACATATGCTTATTCGAACCTACTGGGATTTTCAATTTCTCACC 852
QY 841 ACTCTTACCAATGTTGAGTCTGAGGAGCTCCACTGCAACCTGCAAAACCCCTGACC 900
Db 853 ACTCTTACCAATGTTGAGTCTGAGGAGCTCCACTGCAACCTGCAAAACCCCTGACC 912
QY 901 GAAGAAGTGGATGATCTTCTACAGTGAAGTCTAGGAAGACCCTACGTTATCTGAGAC 960
Db 913 TAAGGAATGGAACACTTTGACAGAGCTCTGGAAGAAATGCTTGGGTTTCTCT 972
QY 961 GGGGTGATGCTGATGACATGACGTCAGAAAGGAGGCAATGTAATGTCATGAGCCCTTGC 1020
Db 973 GGGGTGATGCTGATGACATGACGTCAGAAAGGAGGCAATGTAATGTCATGAGCCCTGAGC 1032
QY 1021 CAAGATCCCAAAAGTCTGAGGAGTCTGAGGAGTCTAGGAAGTCTAGGAAGTCTAGGACT 1080
Db 1033 CAAGATCCCAAAAGTCTGAGGAGTCTGAGGAGTCTAGGAAGTCTAGGAAGTCTAGGACT 1092
QY 1081 CAATACTCGGCTGATCAAGTGGATGATCCCAAGATGATCTTCTGATCCTCCAAAGCAA 1140
Db 1093 CAATACTCGGCTGATCAAGTGGATGATCCCAAGATGATCTTCTGATCCTCCAAAGCAA 1152
QY 1141 AGCTTTATCACTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1200
Db 1153 AGCTTTATCACTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1212
QY 1201 TATGTTGGGAGTTCCTCATTTTGGTGAATGATGATGATGATGATGATGATGATGATGATGAT 1260
Db 1213 TATGTTGGGAGTTCCTCATTTTGGTGAATGATGATGATGATGATGATGATGATGATGATGAT 1272
QY 1261 AGGAGAGCTGTGATGAATTAACCTCAAAAATGATGACAAAGCAAGATTTACTGAGGCTTT 1320
Db 1273 GGGAGAGCTGTGATGAATTAACCTCAAAAATGATGACAAAGCAAGATTTACTGAGGCTTT 1332
QY 1321 GAGAAGAGCTGATTCGATTCCTCTTATTAAGAGATGCTATGAGATTTATCAAGATTTCA 1380
Db 1333 GAGAAGAGCTGATTCGATTCCTCTTATTAAGAGATGCTATGAGATTTATCAAGATTTCA 1392
QY 1381 CCATGATCAACCTGTAAGCCCTAGATCGAGAGCTCTTCTGATCGAGATTTGATGCG 1440
Db 1393 ACATGATCAACCTGTAAGCCCTAGATCGAGAGCTCTTCTGATCGAGATTTGATGCG 1452
QY 1441 CCAGAAAGAGCCCAAGAGCTGCGATCAGCTGCCATGACCTGCTGCTCCAGACTA 1500
Db 1453 CCAGAAAGAGCCCAAGAGCTGCGATCAGCTGCCATGACCTGCTGCTCCAGACTA 1512
QY 1501 CTCTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1560
Db 1513 CTCTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1572
QY 1561 AAAATGTTTATTTTCTGCTGCAAAAATTTAATTAAGATGAGAAAGAGGGA 1620
Db 1573 AAAATGTTTCTGTTTCTGCTGCAAAAATTTAATTAAGATGAGAAAGAGGGA 1632
QY 1621 ATAGATCTTTC 1631
Db 1633 TTAGTTATATC 1643

```

```

RESULT 11
AF401657 1915 bp mRNA linear PRI 11-JUL-2002
LOCUS Macaca fascicularis UDP-glucuronosyltransferase 2B30 (UGT2B30)
DEFINITION
AF401657
ACCESSION AF401657.1 GI:18139799
VERSION

```

```

KEYWORDS
SOURCE Macaca fascicularis (crab-eating macaque)
ORGANISM Macaca fascicularis
Mammalia; Eutheria; Chordata; Craniata; Vertebrata; Euleleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
Cercopithecinae; Macaca.
REFERENCE
1 (bases 1 to 1915)
Giard, C., Bardier, O., Turgeon, D. and Belanger, A.
Isolation and characterization of the monkey UGT2B30 gene that
encodes a uridine diphosphate-glucuronosyltransferase enzyme active
on mineralocorticoid, glucocorticoid, androgen and oestrogen
hormones
Biochem. J. 365 (Pt 1), 213-222 (2002)
MEDLINE 22067079
PUBMED 12071853
REFERENCE
2 (bases 1 to 1915)
Giard, C., Bardier, O. and Belanger, A.
Direct Submission
Submitted (24-JUL-2001) Molecular Endocrinology, CHUL Research
Center, 2705 Laurier Blvd., Sainte-Foy, QC G1V 4G2, Canada
FEATURES
source
1..1915
/organism="Macaca fascicularis"
/mol_type="mRNA"
/db_xref="taxon:9541"
1..1915
/gene="UGT2B30"
1..1587
/gene="UGT2B30"
/codon_start=1
/product="UDP-glucuronosyltransferase 2B30"
/protein_id="AAL60145.1"
/db_xref="GI:18139800"
/translation="MSMKRTSALLILQLSCTSSGNCGRKLVLPFSHMMNKTILD
ELVORGEVAVLAVSPSILPEPNPSPALKKEIPSLTBTEFEDSVTLVKRSDLPK
DTPWPHFLOVQEMMTYGMIRKFCDFVSNKMLKLOSRDVLADLISGCELL
AELIKLPFVYSLEPSPYAIERKGGFLPEPSPVPMSEFSDQMTMEVKKMIYV
YEPFPOAMDTKMDQFSEVLGRPTLTFETMAKAEIWLIRNWDFOFPHLPVEL
VGGHCKPKAPLPKMEBGFVOSGSDGVVAFSGSVSNSEERAVVIALAKTIPOK
VLMRFPENKRPDITGLNTOLYKMLPONDLDGHPRTAPFITHGNALYEALYHGIPIWG
VPLFADQLDIANKAKARAVSLDPNMTSDTLHLAKIVINDPFEKNMKLSIHH
DQVKKPLDRAVFWIEFVMRHKGAKHRLVAAYDLTWQYHSLDIVIGFLCAVAVIFII
TKCLFVCLKFVRGKKGRD"
BASE COUNT 564 a 374 c 404 g 573 t
ORIGIN
Query Match 76.0%; Score 1254.6; DB 9; Length 1915;
Best Local Similarity 87.0%; Pred. No. 3; e-310;
Matches 1390; Conservative 0; Mismatches 204; Indels 3; Gaps 1;
QY 35 ATGCTTATGAATGAGCTTACGCTTCTCTCTGATACGCTGAGCTTACTTAACTCT 94
Db 1 ATGCTTATGAATGAGCTTACGCTTCTCTCTGATACGCTGAGCTTACTTAACTCT 60
QY 95 GGGAGTTGTGAAGAGTGTGTGGGCCACAGAAATTCAGCACTGATGAATATAAG 154
Db 61 GGGATTTGTGAAGAGTGTGTGGGCCACAGAAATTCAGCACTGATGAATATAAG 120
QY 155 ACAATCTGATGAATCTGTCAGAGAGTGCATGAGGTGACTGATTTGATGCTTCAGCT 214
Db 121 ACAATCTGATGAGTGTGTCAGAGAGTGCATGAGGTGACTGATTTGATGCTTCAGCT 180
QY 215 TCCATTTCTTGCATGCCAACAGCCCATCTTAAATTTGAAGTTATCTCTGATCT 274
Db 181 TCCATTTCTTGCATGCCAACAGCCCATCTTAAATTTGAAGTTATCTCTGATCT 240
QY 275 TTAACCTAAAGTGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 334
Db 241 TTAACCTAAAGTGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 300
QY 335 CCAAAAGACACATTTTGGTCATATTTTTCACAGATGACAGAAATCATGTGACATTTAAT 394
Db 301 CCAAAAGACACATTTTGGTCATATTTTTCACAGATGACAGAAATCATGTGACATTTAAT 360

```

QY 395 GACATCTTAGAAGATCTGTAAGGATATAGTTTCAATTAAGAACTTATGAGAACTA 454
 Db 361 GACATGATTAGAAAGTTCTGTAAGATGTGTTTCTAATTAAGAACTTATGAGAACTA 420
 QY 455 CAGAGTCAGAAATTTGATGTTGTTCTTTCAGATGCTGTTTCCCTTGTGAGCTGCTG 514
 Db 421 CAGAGTCAGAAATTTGATGTTGTTCTTTCAGATGCTGTTTCCCTTGTGAGCTGCTG 480
 QY 515 GCCGATTAATTAATACCTTTTGTCTACAGCTTCGCTGCTGCTGCTGCTGCTGCTG 574
 Db 481 GCTGAGCTAATTAATACCTTTTGTCTACAGCTTCGCTGCTGCTGCTGCTGCTGCTG 540
 QY 575 GAAAAGCATAGTGAGGAGCTTCTGTTCCCTCTCTTATGAGCTGTTATGTCAGAA 634
 Db 541 GAAAAGCATAGTGAGGAGCTTCTGTTCCCTCTCTTATGAGCTGTTATGTCAGAA 600
 QY 635 CTAAGTCAGAAATGACTTTCTATAGAGAGGTAATAATATGATCTATGCTTATTTT 694
 Db 601 TTCAGTGAATCAATGACTTTCTATAGAGAGGTAATAATATGATCTATGCTTATTTT 660
 QY 695 GAATTTTGGTTCCAAATATTTGACATGAAAGTGGATCACTTCTACAGTGAAGTTCTA 754
 Db 661 GACTTTTGGTTCCAAAGATGGGATCGAAAGTGGATCACTTCTACAGTGAAGTTCTA 720
 QY 755 GGAAGACCCACCTAGCTTATCTGAGACATGCGAAAGCTGACATAGGCTTATGGAAC 814
 Db 721 GGAAGACCCACCTAGCTTATCTGAGACATGCGAAAGCTGACATAGGCTTATGGAAC 780
 QY 815 TACTGGATTTTCAATTTCTCACCACCTCTTACCAAAATGTTGAGTTGTTGAGAGCTC 874
 Db 781 TACTGGATTTTCAATTTCTCACCACCTCTTACCAAAATGTTGAGTTGTTGAGAGCTC 840
 QY 875 CACTGCAAACTGCGCAAAACCCCTACCGAAAGAAATGAAAGATTTGTCAGAGCTTGA 934
 Db 841 CACTGCAAACTGCGCAAAACCCCTACCGAAAGAAATGAAAGATTTGTCAGAGCTTGA 900
 QY 935 GAAAATGCTGTTGTTGTTTCTGCGGGTGCATGAGTCAATGACATGAGTGAAGAAAG 994
 Db 901 GAAAATGCTGTTGTTGTTTCTGCGGGTGCATGAGTCAATGACATGAGTGAAGAAAG 960
 QY 995 GCCAATGTAATTTGATCAGCCCTTGCAGAGATCCCAAAAGTTCTGTGAGATTTGAT 1054
 Db 961 GCCAATGTAATTTGATCAGCCCTTGCAGAGATCCCAAAAGTTCTGTGAGATTTGAT 1020
 QY 1055 GGGAAATTAACAGATCTTAAAGACTCAATCTCGGCTGTACAGTGAATCCCAAGAT 1114
 Db 1021 GGGAAATTAACAGATCTTAAAGACTCAATCTCGGCTGTACAGTGAATCCCAAGAT 1080
 QY 1115 GATCTTTTGTGATCCCAAAACCAAGCTTTTATCACTCATGAGTGAATGAATGGAGATC 1174
 Db 1081 GATCTTTTGTGATCCCAAAACCAAGCTTTTATCACTCATGAGTGAATGAATGGAGATC 1140
 QY 1175 TATGAAGCTATTTACATGAGGCTTCCCTATGTTGAGAGTTCCCATATTTGATGAGCTT 1234
 Db 1141 TATGAAGCTATTTACATGAGGCTTCCCTATGTTGAGAGTTCCCATATTTGATGAGCTT 1200
 QY 1235 GATTAATAGTCTCATGAGGCTCAAGAGGAGAGCTGTAGAAATTAATCTTCAAACTATG 1294
 Db 1201 GATTAATAGTCTCATGAGGCTCAAGAGGAGAGCTGTAGAAATTAATCTTCAAACTATG 1260
 QY 1295 ACAACGAGATTTTCTGAGGCTTGAAGACATCATACGATTCAGATTCCTTATTAAGAG 1354
 Db 1261 ACAACGAGATTTTCTGAGGCTTGAAGACATCATACGATTCAGATTCCTTATTAAGAG 1320
 QY 1355 AATGCTATGAGTTATCAAGATTCATCATGATCAACTGTAAGGCTTATGAGAGCTTATGAG 1414
 Db 1321 AATGCTATGAGTTATCAAGATTCATCATGATCAACTGTAAGGCTTATGAGAGCTTATGAG 1380
 QY 1415 GTCTTCTGATGAGTTTGTGATGCGCCAAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1474
 Db 1381 GTCTTCTGATGAGTTTGTGATGCGCCAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1440

QY 1475 CATGACCTCACTGTTCCAGACCTACTATATAGATGATGTTGGTTCTGCTGACCTGT 1534
 Db 1441 TAGACCTCACTGTTCCAGACCTACTATATAGATGATGTTGGTTCTGCTGACCTGT 1500
 QY 1535 GTGCAACTGCTATATTTCTTGTACAAATATGTTTATTTTCTGCTGCAAAATTTAT 1594
 Db 1501 GTGCAACTGCTATATTTCTTGTACAAATATGTTTATTTTCTGCTGCAAAATTTAT 1557
 QY 1595 AAAACTGAAAGATAGAAAGAGGAAATGATCTTTT 1631
 Db 1558 AGAAGAGAAAGAGGAAAGAGATTTATATCATC 1594
 RESULT 12
 AX336329
 LOCUS AX336329 1855 bp DNA linear PAT 09-JAN-2002
 DEFINITION Sequence 6838 from Patent WO0194629.
 ACCESSION AX336329
 VERSION AX336329.1 GI:18127048
 KEYWORDS
 SOURCE
 ORGANISM
 Homo sapiens (human)
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 FEATURES
 source
 1. 1855
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"
 location/Qualifiers
 BASE COUNT 571 a 337 c 394 g 553 t
 ORIGIN
 Query Match 75.9%; Score 1252.6; DB 6; Length 1855;
 Beat Local Similarity 86.1%; Pred. No. 1.2e-309;
 Matches 1387; Conservative 0; Mismatches 224; Indels 0; Gaps 0;
 QY 21 TGCAATGATGAGATGCTATGAAATGAGACTTCACTTCTCTGATAGAGCTGACT 80
 Db 1 TGCAATGATGAGATGCTATGAAATGAGACTTCACTTCTCTGATAGAGCTGACT 60
 QY 81 GTTACTTAACTCTGAGAGTTGTGAAAGGTGTGTGTCGCCACAGAAATTCAGCCACT 140
 Db 61 TTTGCTTAACTCTGAGAGTTGTGAAAGGTGTGTGTCGCCACAGAAATTCAGCCACT 120
 QY 141 GATGAATATTAAGACATCTCTGATGAACTTGTCCAGAGAGTATAGAGGATGACTGAT 200
 Db 121 GATGAATATTAAGACATCTCTGATGAACTTGTCCAGAGAGTATAGAGGATGACTGAT 180
 QY 201 TGGCATCTTCACTGCTTCAATTTCTTGCATCCCAAGGCCATCTTCTTAAATTTGAG 260
 Db 181 TGGCATCTTCACTGCTTCAATTTCTTGCATCCCAAGGCCATCTTCTTAAATTTGAG 240
 QY 261 TTTATCTGATATCTTAACTAAAGATTTGAGATTTATCAAGAGCTGTTAAGA 320
 Db 241 TTTATCTGATATCTTAACTAAAGATTTGAGATTTATCAAGAGCTGTTAAGA 300
 QY 321 GATGGCAGAACTTCCAAAGACATTTTGTGCTATTTTTCACAAGTACAGAAATCA 380
 Db 301 GATGGCAGAACTTCCAAAGACATTTTGTGCTATTTTTCACAAGTACAGAAATCA 360
 QY 381 TGTGCAATTTAATGACATCTTAAAGATTTCTGTAAGATATAGTTTCAATTAAGAAC 440
 Db 361 TGTGCAATTTAATGACATCTTAAAGATTTCTGTAAGATATAGTTTCAATTAAGAAC 420
 QY 441 TTATGAGAAATCAAGAGTCAAGATTTGATGTTGTTCTTGCAGATGCTGTTTCCCT 500

```

Db 421 TTATGAAAAAGTACAGAGTCAAGATTGACGTATTTTTCAGATGCTATTTTCCCT 480
Qy 501 TTGGTAGCTCTGAGCCGAGTTACTTAAATACCTTTGTCTACAGCCTCGCTTCTC 560
Db 481 GTAGTAGCTCTGAGCCGAGTTACTTAAATACCTTTGTCTACAGCCTCGCTTCTC 540
Qy 561 CTGGCTACGCAATTGAAAAAGCATAGTGAAGACTTCTGTTCCCTTCTCTATGTCCTG 620
Db 541 CTGGCTACGCAATTGAAAAAGCATAGTGAAGACTTCTGTTCCCTTCTCTATGTCCTG 600
Qy 621 TTGTTATGTCAGAACTAAGTACCAATGACTTTCTATGAGAGGCTAAAAATATGATCT 680
Db 601 TTGTTATGTCAGAACTAAGTACCAATGACTTTCTATGAGAGGCTAAAAATATGATCT 660
Qy 681 ATGCTTATTTGAAATTTGTTCCAAATTTGACATGAAGAAGGGAGTACGTTCT 740
Db 661 ATGCTTATTTGAAATTTGTTCCAAATTTGACATGAAGAAGGGAGTACGTTCT 720
Qy 741 ACAGTGAAGTTCTAGAGAACCCACTACGTTATCTGAGACAAATGSCAAAAGCTGACATAT 800
Db 721 ATAGTGAAGTTCTAGAGAACCCACTACGTTATCTGAGACAAATGSCAAAAGCTGACATAT 780
Qy 801 GCGTTATTCGAATCTAGGAGTTTTCATTTCTCACCCACTTTACCAATGTTGAGT 860
Db 781 GCGTTATTCGAATCTAGGAGTTTTCATTTCTCACCCACTTTACCAATGTTGAGT 840
Qy 861 TCGTTGAGAGCTCAGTCAAGACCTGCAACCCCTACGAAAGGAAATGGAAGATTG 920
Db 841 TTGTTGAGAGCTCAGTCAAGACCCCTGCAACCCCTGCTAAGAAATGGAAGATTG 900
Qy 921 TCCAGAGCTCTGAGAAATGCTGTGTGTGTGTTTCTGAGGGTCGATGTGCTAGTAACA 980
Db 901 TAGAGAGCTCTGAGAAATGCTGTGTGTGTGTTTCTGAGGGTCGATGTGCTAGTAACA 960
Qy 981 CGTGAAGAAAGGCGCCATGTAATGCTAGCCTTCCAGAGTCCACAAAAGTTTC 1040
Db 961 TGACAGAAAGGCGCCATGTAATGCTAGCCTTCCAGAGTCCACAAAAGTTTC 1020
Qy 1041 TGTCGAGTTTGTAGGGAAATTAACCAATCTTGAAGCTCAATCTCGGTGTCAGTAAGT 1100
Db 1021 TGTCGAGTTTGTAGGGAAATTAACCAATCTTGAAGCTCAATCTCGGTGTCAGTAAGT 1080
Qy 1101 GGATACCCCAAGATGATCTTCTGTGCTATCCCAAAACCAAGCTTTTATCACTGATG 1160
Db 1081 GGATACCCCAAGATGATCTTCTGTGCTATCCCAAAACCAAGCTTTTATCACTGATG 1140
Qy 1161 GAATGAATGGATGTATGAAGCTATTAACATGAGGCTTCTATGCTGAGGATCCCATAT 1220
Db 1141 GAGCCATGGATGTATGAAGCTATTAACATGAGGCTTCTATGCTGAGGATCCCATAT 1200
Qy 1221 TTGGTATCAGCTTGTATACCTACTACATGAAGGCGCAAGAGCAGCTGTAGAAATTA 1280
Db 1201 TTGGTATCAGCTTGTATACCTACTACATGAAGGCGCAAGAGCAGCTGTAGAAATTA 1260
Qy 1281 ACTTCAAAAATATGACAAAGCAAGATTTTACGAGGCTTTGAGAAACAGTCACTTACGAT 1340
Db 1261 ACTTCAAAAATATGACAAAGCAAGATTTTACGAGGCTTTGAGAAACAGTCACTTACGAT 1320
Qy 1341 CCTCTTAATAAGAGATGCTATGAAGATTAACAGAAATTCACCATGATCAACCTGTAAGC 1400
Db 1321 CCTCTTAATAAGAGATGCTATGAAGATTAACAGAAATTCACCATGATCAACCTGTAAGC 1380
Qy 1401 CCTCTTAATAAGAGATGCTATGAAGATTAACAGAAATTCACCATGATCAACCTGTAAGC 1460
Db 1381 CCTCTTAATAAGAGATGCTATGAAGATTAACAGAAATTCACCATGATCAACCTGTAAGC 1440
Qy 1461 TGCGATAGCTGCGCCATGACCTTCACTGCTTCCAGACTACTATATGATGTTGGGT 1520
Db 1441 TTGGGTTGAGCCCAAGACCTCACTGCTTCCAGACTACTATATGATGTTGGGT 1500
Qy 1521 TCTGCTGACCTGTGTGGCAACTGCTATATCTTGTTCACAAAATGTTTTTATTTTCT 1580
Db 1501 TCTGCTGACCTGTGTGGCAACTGCTATATCTTGTTCACAAAATGTTTTTATTTTCT 1560

```

```

Qy 1581 GTCAAAAATTTAATTAATACTAGAAAAGATAGAAAAGAGGATAGATCTTC 1631
Db 1561 TCTGGAAGTTTGTCTAGAAAAGCAAGAGAGGAAAAAATGATTAATATATC 1611

RESULT 13
AX336696 1855 bp DNA linear PAT 09-JAN-2002
LOCUS Sequence 7205 from Patent WO0194629.
DEFINITION AX336696
ACCESSION AX336696
VERSION AX336696.1 GI:18127415
KEYWORDS
SOURCE
ORGANISM Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 Young, P.E., Augustus, M., Carter, K.C., Ebner, R., Endress, G.,
Horrihan, S., Soppet, D.R. and Weaver, Z.
Cancer gene determination and therapeutic screening using signature
gene sets
Patent: WO 0194629-A 7205 13-DEC-2001;
Avalon Pharmaceuticals (US)
Location/Qualifiers
FEATURES
source
1..1855
/mol_type="genomic DNA"
/db_xref="taxon:9606"

BASE COUNT 571 a 337 c 394 g 553 t
ORIGIN

Query Match 75.9%; Score 1252.6; DB 6; Length 1855;
Best Local Similarity 86.1%; Pred. No. 1.2e-309;
Matches 1387; Conservative 0; Mismatches 224; Indels 0; Gaps 0;

Qy 21 TGCAATGATCAGAGATGCTATGAAATGAGACTTCACTCTTCTCTGATACAGTGAAGT 80
Db 1 TGCAATGATCAGAGATGCTATGAAATGAGACTTCACTCTTCTCTGATACAGTGAAGT 60
Qy 81 GTTACTTACCTCTGAGAGTGTGAAAGGCTGTGTGTGAGGCGCCACAGAAATTCAGCCACT 140
Db 61 TTGCTTTACCTCTGAGAGATGTGAAAGGCTGTGTGTGAGGCGCCACAGAAATTCAGCCACT 120
Qy 141 GGAATGAATTAAGACAACTCTGAGATGAACTTCTCCAGAGAGGTCATGAGAGTACTAT 200
Db 121 GGAATGAATTAAGACAACTCTGAGATGAACTTCTCCAGAGAGGTCATGAGAGTACTAT 180
Qy 201 TGCAATCTTACGCTTCAATTTCTTGATGCCAACAGCCCATCTACTCTTAAATTTGAG 260
Db 181 TGCAATCTTACGCTTCAATTTCTTGATGCCAACAGCCCATCTACTCTTAAATTTGAG 240
Qy 261 TTTATCTGTATCTTTAATACTAAACGAGTTTGAGATATTTATCAAGCAGCTGGTTAGA 320
Db 241 TTTATCTGTATCTTTAATACTAAACGAGTTTGAGATATTTATCAAGCAGCTGGTTAGA 300
Qy 321 GATGGGAGAACTTCCAAAAGACATTTTGTGCTATTTTTCACAAAGTACAAAGAAATCA 380
Db 301 GATGGGAGAACTTCCAAAAGACATTTTGTGCTATTTTTCACAAAGTACAAAGAAATCA 360
Qy 381 TGTCGACATTATTAAGACATCTTGAAGAGTTCTGTAAGATATAGTTTCAATTAAGAAC 440
Db 361 TGTCGACATTATTAAGACATCTTGAAGAGTTCTGTAAGATATAGTTTCAATTAAGAAC 420
Qy 441 TTATGAAGAACTACAGAGTCAAGATTTGATGTTCTTTCAGAGATGCTGTTTCCCT 500
Db 421 TTATGAAGAACTACAGAGTCAAGATTTGATGTTCTTTCAGAGATGCTGTTTCCCT 480
Qy 501 TTGGTAGCTCTGAGCCGAGTTACTTAAATACCTTTGTCTACAGCCTCGCTTCTC 560
Db 481 GTAGTAGCTCTGAGCCGAGTTACTTAAATACCTTTGTCTACAGCCTCGCTTCTC 540
Qy 561 CTGGCTACGCAATTGAAAAAGCATAGTGAAGACTTCTGTTCCCTTCTCTATGTCCTG 620

```

| | | | |
|----|------|---|------|
| Db | 541 | CTGGCTACACTTTTGGAAAAAGCATAGTGGAGGATTTATTTTCCCTCTCTCTACGTAACCTGG | 600 |
| Oy | 621 | TTGTATATGTCAGAACTAAGTAGACCAAAATGACTTTTCATAGAGGGGTAAAAAATAGATCT | 680 |
| Db | 601 | TTGTATATGCAAGAAATTAACTGATCAAAATGACTTTTCATGAGAGGGGTAAAAAATAGATCT | 660 |
| Oy | 681 | ATGTCCTTTATTTTGAATTTTGGTTCCAAATATTTTGACATGAAGAAGTGGATCTGCTTC | 740 |
| Db | 661 | ATGTCCTTTACTTTTGAATTTTGGTTCGAAATTTTTCGACATGAAGAAGTGGATCTGCTTC | 720 |
| Oy | 741 | AACGGAAGTCTTAGAAGAACCCATACGTTATCTGAGCAATGGCAAAAAGCTGACATAT | 800 |
| Db | 721 | ATAGTGAAGTCTTAGAAGAACCCATACGTTATCTGAGCAATGGGGAAGCTGACATAT | 780 |
| Oy | 801 | GGCTTATTCGAAACTACTGGGATTTTTCAAATTTCTCACCACACTTTACCAATGTTTGAAT | 860 |
| Db | 781 | GGCTTATTCGAAACTCTGGGATTTTTCAGTTTCTCCTCATCTTACCAATGTTTGAAT | 840 |
| Oy | 861 | TCGTTGAGAGACTCCAATGCAAACTGCCAAACCCCTTACCGAAGGAATGAAAGCTTGG | 920 |
| Db | 841 | TTGTTGGAGAGACTCCAATGCAAACTGCCAAACCCCTGCTTAAGGAATGAAAGACTTGG | 900 |
| Oy | 921 | TCCAGAGCTCTGGAGAAAAATGGGTGTGGGTGTTTCTCTGGGGTCGATGGTCAGTAACA | 980 |
| Db | 901 | TACAGAGCTCTGGAGAAAAATGGGTGTGGGTGTTTCTCTGGGGTCGATGGTCAGTAACA | 960 |
| Oy | 981 | CGTCAGAAAGAAAGGGCCCAATGTAAATTTGCAATCAGCCCTTGGCCCAAGATCCCACAAAAGTTTC | 1040 |
| Db | 961 | TGACAGAAAGAAAGGGCCCAATGTAAATTTGCAATCAGCCCTGCGCCAGATCCCACAAAAGTTTC | 1020 |
| Oy | 1041 | TGTGAGATTTGATGGGAATTAACCAAGATATCTTTAGAGTCAATATCTGGCTGTACAAGT | 1100 |
| Db | 1021 | TGTGAGATTTGATGGGAATTAACCAAGATATCTTTAGAGTCAATATCTGGCTGTATAAGT | 1080 |
| Oy | 1101 | GGATACCCCAAGATATATCTTTGGTGCATCCCAAAACCAACCTTTATACCTCATAGGTG | 1160 |
| Db | 1081 | GGATACCCCAAGATATATCTTTAGGTGCATCCCAAAACCAACCTTTTATATCTCATAGGTG | 1140 |
| Oy | 1161 | GAATGAATGGGATCTATGAGCTTATTTACATGGGGTCCCTATGTTGGGAGTTCCCATAT | 1220 |
| Db | 1141 | GAGCCAAATGGCATCTATGAGGCAATCTACCATGGGATCCCTATGTTGGGAGTTCCCATGCT | 1200 |
| Oy | 1221 | TTGTGATCAAGCTTGTATTAACATAGCTCACATGAAGGCCAAAGAGCAAGCTGTAGAAATTA | 1280 |
| Db | 1201 | TTGCGGATCAACCTGTATTAACATTTGCTCATAGTAAGGCCAGGGAGAGCAAGCTGTAGAGTGG | 1260 |
| Oy | 1281 | ACTTAAAACTATGACAGCGAAGATTTATCTGAGGCTTTTGAGAACATGATATTCGATAT | 1340 |
| Db | 1261 | ACTTAAACAATGTCAGATACAGACTTGTCTAATGCAATTAAGAAAGATTAATATGATC | 1320 |
| Oy | 1341 | CCTCTTATTAAGAGATGCTATGAGATTTATCAAGAAATTCACCATGATCAACCTGTAAAGC | 1400 |
| Db | 1321 | CTTCAATATTAAGAGATGTTATGAATTTATCAAGAAATTCACCATGATCAACCATGAGAC | 1380 |
| Oy | 1401 | CCCTAGATGAGCAGTCTTTCTGAGATGAGTTTGTCAATGGGCCAACAAGAGGCCAAGCACCC | 1460 |
| Db | 1381 | CCCTGAGATGAGCAGTCTTTCTGAGATGAGTTTGTCAATGGGCCAACAAGAGGCCAACAACC | 1440 |
| Oy | 1461 | TGCGATCAGCTGCCCATGACCTTCACTGGTTCAGCATCTATCTATAGATGTGATTTGGGT | 1520 |
| Db | 1441 | TTGCGGATGACACCCCAAGCACTTCACTGGTTCAGCATCTTTTGATGTGATTTGGGT | 1500 |
| Oy | 1521 | TCCTGCTGACGTTGTGGCAATCGCATATTTCTGTGTCACAAAATGTTTTTATTTATTTCCCT | 1580 |
| Db | 1501 | TCCTCTGCTGTGTGTGGCAATCGTGATATTTATCTCTCAAAAATGTTTGTCTGTTTTTGTTC | 1560 |
| Oy | 1581 | GTCAAAAAATTTATATAAACTAGAAAGATGAAAAAGAGGGAATAGATCTTTC | 1631 |
| Db | 1561 | TCTGAAAGTTTCTAGAAAAAGCAAGGAAGGAATAATTAATTAATATC | 1611 |

| AX09473 | AX09473 | 1855 bp | DNA | linear | PAT 14-JUN-2002 |
|-----------------------|--------------|---|-----------------|--------------|-----------------|
| LOCUS | AX09473 | Sequence 2120 from Patent WO0229103. | | | |
| DEFINITION | AX09473 | | | | |
| ACCESSION | AX09473 | | | | |
| VERSION | AX09473.1 | GI:21442178 | | | |
| KEYWORDS | | | | | |
| SOURCE | | | | | |
| ORGANISM | | Homo sapiens (human)- | | | |
| REFERENCE | | Homo sapiens | | | |
| AUTHORS | | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; | | | |
| TITLE | | Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. | | | |
| JOURNAL | | Alvares, C., Horne, D., Pires-da-Silva, S. and Vockley, J.G. | | | |
| FEATURES | | Gene expression profiles in liver cancer | | | |
| source | | Patent: WO 0229103-A 2120 11-APR-2002; | | | |
| | | GENE LOGIC INC (US) | | | |
| | | Location/Qualifiers | | | |
| | | 1..1855 | | | |
| | | /organism="Homo sapiens" | | | |
| | | /mol_type="genomic DNA" | | | |
| | | /db_xref="taxon:9606" | | | |
| | | /note="EMBL/Genbank Accession No. J05428" | | | |
| BASE COUNT | 571 a | 337 c | 394 g | 553 t | |
| ORIGIN | | | | | |
| Query Match | 75.9% | Pred. 1252.6; | DB 6; | Length 1855; | |
| Best Local Similarity | 86.1% | Fold. No. 1.2e-309; | | | |
| Matches 1387; | Conservative | 0; | Mismatches 224; | Indels 0; | Gaps 0. |
| QY | 21 | TGCATTGATGAGATGTCATGAATGAGCTTCCTCTCGATACAGCTGACCT | 80 | | |
| DB | 1 | TGCATTGACACAGATGTCGTGAATGACTTCAGTAATTTGGTATACACTGACCT | 60 | | |
| QY | 81 | GTTACTTAACTCTGGAGATTGGAAAGTGCTGTGTGGCCACAGAAATTCAGCCACT | 140 | | |
| DB | 61 | TTTGGTTAGCTCTGGGAATGTGGAAAGTGCTGTGTGGCCACAGAAATTCAGCCACT | 120 | | |
| QY | 141 | GGATGAATATAAGCATCTCGATGAGCTTCGACAGAGAGTCATGAGGACTGAT | 200 | | |
| DB | 121 | GGATGAATATAAGCATCTCGATGAGCTTCGACAGAGAGTCATGAGGACTGAT | 180 | | |
| QY | 201 | TGGCATCTTCAGCTCCATCTTCTTCGATCCCAACAGCCCATCTTAAATTGGAAG | 260 | | |
| DB | 181 | TGGCATCTTCAGCTCCATCTTCTTCGATCCCAACAGCCCATCTTAAATTGGAAG | 240 | | |
| QY | 261 | TTTATCTGTATCTTTAACTAAACCTGAGTTGAGATATTAACAAGCAGCTGTTAAGA | 320 | | |
| DB | 241 | TTTATCTGTATCTTTAACTAAACCTGAGTTGAGATATTAACAAGCAGCTGTTAAGA | 300 | | |
| QY | 321 | GATGGGCAAGCTCCCAAGACATTTTGTCTATTTTTCACAAGTCAAGAAATCA | 380 | | |
| DB | 301 | GATGGGCAAGCTCCCAAGACATTTTGTCTATTTTTCACAAGTCAAGAAATCA | 360 | | |
| QY | 361 | TGTGACATTTAATGACATCTTCTGAAAGTCTGTGAAGATATAGTTCAATAAAGAAC | 440 | | |
| DB | 361 | TGTGACATTTAATGACATCTTCTGAAAGTCTGTGAAGATATAGTTCAATAAAGAAC | 420 | | |
| QY | 441 | TTATGAGAACTACAGAGTCAAGATTTGATGTTCTTGACAGATCTGTTTCCCT | 500 | | |
| DB | 421 | TTATGAGAACTACAGAGTCAAGATTTGATGTTCTTGACAGATCTGTTTCCCT | 480 | | |
| QY | 501 | TTGTGAGAGCTGTGGCCGAGTTACTTAAATATCCCTTTGTCTACAGCTCCGCTCTCTC | 560 | | |
| DB | 481 | TTGTGAGAGCTGTGGCCGAGTTACTTAAATATCCCTTTGTCTACAGCTCCGCTCTCTC | 540 | | |
| QY | 561 | CTGGCTAGGCAATTGAAAGCATAGTAGAGACTCTGTGCTCCCTCTCTCTATGTGCTG | 620 | | |
| DB | 541 | CTGGCTAGGCAATTGAAAGCATAGTAGAGACTCTGTGCTCCCTCTCTCTATGTGCTG | 600 | | |
| QY | 621 | TTGTATGTCAAGATTAAGTACCAATATGATCTTTTATGAGAGAGGTTAAAAATATGATCT | 680 | | |
| DB | 601 | TTGTATGTCAAGATTAAGTACCAATATGATCTTTTATGAGAGAGGTTAAAAATATGATCT | 660 | | |
| QY | 681 | ATGTCCTTATTTTGAATTTTGGTTCCAAATATTTTGAATGAAGAAGTGAGATCAGTCT | 740 | | |

| | | | |
|----|------|--|------|
| Db | 661 | ATGTCCTTACTTTGACCTTTTGTTGGTTCGAAATATTTTGACATGAAAGTGGATCAGTTT | 720 |
| Qy | 741 | ACAGTGAAGTCTTAGAAGAACCCACCTACGTTATCTGAGACAAATGGCAAAAGCTGACATAT | 800 |
| Db | 721 | ATAGTGAAGTCTTAGAAGAACCCACCTACGTTATCTGAGACAAATGGGAAAGCTGACGAT | 780 |
| Qy | 801 | GGCTTATTCGAAACTACTGCGGATTTTCAATTTCTCACCCTCTTACCAATGTTGAGT | 860 |
| Db | 781 | GGCTTATTCGAAACTCTGGAATTTTCAGTTTCTCATTCACCTCTTACCAATGTTGAT | 840 |
| Qy | 861 | TCGTTGAGGAGCTCCACTGCAAACTGCGCAAACTCCCTACCGAGGAATGGAAGTTTG | 920 |
| Db | 841 | TTGTTGAGGAGCTCCACTGCAAACTGCGCAAACTCCCTCTTAAAGGAATGGAAGCTTGG | 900 |
| Qy | 921 | TCCAGAGCTCTGAGAAATGATGTTGTTGAGTATTTTCTCTGGGGTCGATGCGCACTAAC | 980 |
| Db | 901 | TACAGAGCTCTGAGAAATGATGTTGTTGAGTATTTTCTCTGGGGTCAAATGCGCACTAAC | 960 |
| Qy | 981 | CGTCAGAGAAAGGGCCCAATGTAATTTGATCAGCCCTTGCCAAAGATCCCAAAAGTTTC | 1040 |
| Db | 961 | TGACAGAGAAAGGGCCCAAGTAATTTGATCAGCCCTTGCCCAAGATCCCAAAAGTTTC | 1020 |
| Qy | 1041 | TGTGAGATTTGATGCGGAATTAACCAATCTTTTGGACTCAATCTGGCTGTCAAGT | 1100 |
| Db | 1021 | TGTGAGATTTGATGCGGAATTAACCAAGATCCTTTGGTCTCAATCTGGCTGTATAAGT | 1080 |
| Qy | 1101 | GGATTAACCCAGATGATCTCTTGCTGTCATCCCAAAACCAAGGCTTTTATCCTCATGATG | 1160 |
| Db | 1081 | GGATTAACCCAGATGATCCTTTAGGTGTCATCCAAACCAAGGCTTTTATCCTCATGATG | 1140 |
| Qy | 1161 | GAATGAATGGGATCTATGAGCTATTTTACCATGGGGTCCCTATGSGTGGAGATGCCATAT | 1220 |
| Db | 1141 | GAGCCAAATGGCATCTACGAGGCAATCTACATGSGATCCCTATGSGTGGAGATTCATTTGT | 1200 |
| Qy | 1221 | TTGTGATCAGCTTGATTAACAATAGCTCAATGAAAGGCCAAAGAGCAGCTGTGAATAA | 1280 |
| Db | 1201 | TTGCCGATCAACCTGATTAACAATGTGCTCAATGAAAGGCCAGGAGCAGCTGTGAATG | 1260 |
| Qy | 1281 | ACTTCAAAACTATGACAAAGCGAAGTTTACTGAGGGCTTTGAAACAGTCAATACCGATT | 1340 |
| Db | 1261 | ACTTCAACACATGTGCGATACAGACTTGCTGAAATGCAATGGAAGATTAATTAATGATC | 1320 |
| Qy | 1341 | CCTCTTAATAAGAGATGCTATGATATATCAAGAAATTCACCATGATCAACCTGTAAGC | 1400 |
| Db | 1321 | CTTCAATATAAGAGATGTTATGAATATTAACAAGATTAACAATGATCAACAGGAAGC | 1380 |
| Qy | 1401 | CCCTAGATCGAGCAGTCTTCTGSAATCGAGTTTGTATCGGCCCAAAAGAGCAGACACC | 1460 |
| Db | 1381 | CCCTGAGATCGAGCAGTCTTCTGSAATTTTGTATCGGCCCAAAAGAGATTAACACC | 1440 |
| Qy | 1461 | TGCGATCAGCTGCCCATGACCTCACCTGTTCCAGACACTACTCTATAGATGATTTGGT | 1520 |
| Db | 1441 | TTGCGGTTGACAGCCACGACCTCACCTGTTCCAGATCAACCTTTTGATGATTTGGT | 1500 |
| Qy | 1521 | TCTCTGCTGACCTGTGTGCGCACTGCTAATTTCTTGTCACAAAATGTTTTTATTTTCT | 1580 |
| Db | 1501 | TCTCTGCTGCTGTGTGTGCGCACTGCTAATTTCTGTCACAAAATGTTGTCTGTTTGT | 1560 |
| Qy | 1581 | GTCAAAAATTTAATTAACCTGAAGATGAGAAAGAGGATAGATCTTTC | 1631 |
| Db | 1561 | TCTGGAATTTGCTGAGAAAGCAAAAGGGAATAATGATTAATATC | 1611 |

| | |
|------------|--|
| RESULT | 15 |
| HUMUDPETA | |
| LOCUS | |
| DEFINITION | HUMAN 3,4-CATECHOL ESTROGEN UDP-GLUCURONYLTRANSFERASE mRNA, complete cds. |
| ACCESSION | J05428 |
| VERSION | J05428.1 GI:340079 |
| KEYWORDS | 3,4-catechol estrogen UDP-glucuronosyltransferase. Homo sapiens (human) |
| SOURCE | |

[illegible]

QY 501 TTGGAGACCTGCTGGCCGAGTTACTTAATAATACCTTTGTCTACAGACCTCCGCTCTCTC 560
Db 481 GTAGTGAAGCTGCTGCTGAGCTATTTAACTATCCCTTTGTGTCTGCTGCTCTCTCTC 540
QY 561 CTGGCTAGCAATTTGAAAGCAATAGTGGAGAGCTTCTGTTCCCTCTCTCTCTGCTG 620
Db 541 CTGGCTACCTTTGAAAGCAATAGTGGAGAGTTATTTCTCTCTCTCTCTCTG 600
QY 621 TTGTTATCTGCAATAGTACCAATGCACTTTTCATGAGAGGTTAAATAATATGATCT 680
Db 601 TTGTTATCTGCAATTTAACTGATCAATGACTTTTCATGAGAGGTTAAATAATATGATCT 660
QY 681 ATGTGCTTATTTGATTTGTTGTTCCAAATATTTGACATGAGAGTGGATCAGTTCT 740
Db 661 ATGTGCTTATCTTTGATCTTTGTTGTTCCAAATATTTGACATGAGAGTGGATCAGTTCT 720
QY 741 ACAGTGAAGTTCTAGGAAGACCCACTAGCTTATCTGAGCAATGAGCAAAAGCTGACATAT 800
Db 721 ATAGTGAAGTTCTAGGAAGACCCACTAGCTTATCTGAGCAATGAGCAAAAGCTGACATAT 780
QY 801 GGTCTTATCGAAACTACTGAGATTTTCAATTTCCCTCACCCACTTTACCAATGTTGAGT 860
Db 781 GGTCTTATCGAAACTACTGAGATTTTCAATTTCCCTCACCTTCTTACCAATGTTGAGT 840
QY 861 TCGTTGAGAGACTCCACTGCAAACTGCAAAACCCCTACCGAAGAGAAATGGAAGATTG 920
Db 841 TTGTTGAGAGACTCCACTGCAAACTGCAAAACCCCTGCTAAGAGAAATGGAAGACTTG 900
QY 921 TCCAGAGCTCTGAGAAAAATGTTGTTGTTGTTTCTCTGAGGTCGATGTCAGTAACA 980
Db 901 TACAGAGCTCTGAGAAAAATGTTGTTGTTGTTTCTCTGAGGTCGATGTCAGTAACA 960
QY 981 CGTCAAGAAAGAGGCGCAATGTAATTTGATCAGCCCTTGCCAAAGATCCCAAAAGTTTC 1040
Db 961 TGACAGAAAGAGGCGCAATGTAATTTGATCAGCCCTTGCCAAAGATCCCAAAAGTTTC 1020
QY 1041 TGTGAGATTTGATGGAATTAACAGATACTTTAGACTCAATCTCGGCTGTACAAGT 1100
Db 1021 TGTGAGATTTGATGGAATTAACAGATACTTTAGACTCAATCTCGGCTGTACAAGT 1080
QY 1101 GGATACCCGAGATGATCTTCTGTTGATCCCAAAACCAAGCTTTTATCACTCATGTTG 1160
Db 1081 GGATACCCGAGATGATCTTCTGTTGATCCCAAAACCAAGCTTTTATCACTCATGTTG 1140
QY 1161 GAATGAATGGATCTATGAGAGTATTTACATGGGGTCCCTATGTTGGAGTTCCCATAT 1220
Db 1141 GAGCCAAATGGATCTATGAGAGTATTTACATGGGGTCCCTATGTTGGAGTTCCCATAT 1200
QY 1221 TTGATGATGAGCTTATTAATAGCTCAATGAGGCAAGGAGAGAGCTGTAGAAATTA 1280
Db 1201 TTGATGATGAGCTTATTAATAGCTCAATGAGGCAAGGAGAGAGCTGTAGAAATTA 1260
QY 1281 ACTTCAAACTATGACAAAGCAAGATTTACTGAGGGCTTTGAGAACAGTCAATTAACGAT 1340
Db 1261 ACTTCAAACTATGACAAAGCAAGATTTACTGAGGGCTTTGAGAACAGTCAATTAACGAT 1320
QY 1341 CCTCTTATTAAGAGATGCTATGAGATTTACAGAAATTTACCATGATCACTGTAAAGC 1400
Db 1321 CCTCTTATTAAGAGATGCTATGAGATTTACAGAAATTTACCATGATCACTGTAAAGC 1380
QY 1401 CCTGATGAGAGATCTTCTGATGAGTTGTCATGGGCCCAAAAGGAGGCAAGCACC 1460
Db 1381 CCTGATGAGAGATCTTCTGATGAGTTGTCATGGGCCCAAAAGGAGGCTAAACACC 1440
QY 1461 TGGGATCAAGCTCCCATGACCTCACTGTTCCAGACACTACTCTATAGATGATGGGT 1520
Db 1441 TGGGATGAGAGCTCCCATGACCTCACTGTTCCAGACACTCTTGTGATGATGGGT 1500
QY 1521 TCTGCTGACCTGTTGGGCACTGCTATATTTCTGTTCAAAATGTTTTTATTTTCT 1580
Db 1501 TCTGCTGCTGTTGGGCACTGCTATATTTTATGCTCAAAATGTTTTTCTGTTTCT 1560
QY 1581 GTCAAAATTTATTAATACTAGAAAGATGAAAGAGGAGATGATCTTTC 1631

Db 1561 TCTGGAAGTTGCTAGAAAGCAAAAGGAGGAAAAATGATTAATATATC 1611

Search completed: December 7, 2003, 06:56:33
Job time : 4110 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 7, 2003, 03:51:27 ; Search time 349 Seconds
(without alignments)
12762.366 Million cell updates/sec

Title: US-09-980-729B-10

Perfect score: 1650
Sequence: 1 agcacctgcgaagaacagcat.....ccaattccaagaagacctg 1650

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : N Geneseq 19Jun03.*
1: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.*
2: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.*
3: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.*
4: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.*
5: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.*
6: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.*
7: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.*
8: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.*
9: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.*
10: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.*
11: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.*
12: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.*
13: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.*
14: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.*
15: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.*
16: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.*
17: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.*
18: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.*
19: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.*
20: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.*
21: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.*
22: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.*
23: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT.*
24: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.*
25: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2003.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|--------|-------------|--------|----------|--------------------|
| 1 | 1650 | 100.0 | 21 | AAC65396 | Human carbohydrate |
| 2 | 1445.4 | 87.6 | 2092 | AAZ95199 | Human UDP-glucuron |
| 3 | 1437.6 | 86.8 | 2093 | ABN97347 | Gene #3345 used to |
| 4 | 1428 | 86.5 | 2111 | AA569712 | DNA encoding novel |
| 5 | 1331.4 | 80.7 | 1639 | AAAL4190 | Drug metabolising |
| 6 | 1272.6 | 77.1 | 1991 | AAD45991 | Human UGT2B7 exon |
| 7 | 1252.6 | 75.9 | 1855 | ABN95622 | Gene #2120 used to |
| 8 | 1252.6 | 75.9 | 24 | ABL68501 | Kidney cancer rela |

| | | | | | | |
|----|--------|------|------|----|----------|---------------------|
| 9 | 1252.6 | 75.9 | 1855 | 24 | ABL68868 | Kidney cancer rela |
| 10 | 1247.8 | 75.6 | 1854 | 21 | AAZ95200 | Human UDP-glucuron |
| 11 | 1238 | 75.0 | 1722 | 22 | AAD06860 | Human breast cancer |
| 12 | 1235.2 | 74.3 | 2799 | 24 | ABN97261 | Gene #3759 used to |
| 13 | 1221.4 | 74.0 | 1713 | 22 | AAD06844 | Human breast cancer |
| 14 | 1207 | 73.2 | 1859 | 23 | AA569710 | DNA encoding novel |
| 15 | 1199.4 | 72.7 | 2802 | 23 | AA569711 | Uridine diphospho- |
| 16 | 1159.6 | 70.3 | 2107 | 19 | AAV15900 | Human UDP-glucuron |
| 17 | 1126.8 | 68.3 | 1976 | 21 | AAZ95206 | Human CDNA differe |
| 18 | 1125.2 | 68.2 | 2090 | 24 | ABK84210 | Human CDNA |
| 19 | 1125.2 | 68.2 | 2090 | 24 | ABN96795 | Gene #3293 used to |
| 20 | 1125.2 | 68.2 | 2090 | 25 | ABT17080 | Androgen-independe |
| 21 | 1115 | 67.6 | 2123 | 24 | ABN96789 | Gene #3287 used to |
| 22 | 967.4 | 58.6 | 1636 | 24 | AAD24666 | Human drug metabol |
| 23 | 967.4 | 58.6 | 1705 | 25 | ABX72195 | Human NOVX polynuc |
| 24 | 967.4 | 58.6 | 2759 | 24 | ABK12422 | CDNA encoding huma |
| 25 | 967.4 | 58.6 | 2974 | 22 | AA546185 | Human DNA encoding |
| 26 | 967.4 | 58.6 | 2974 | 25 | ACA57943 | Human PRO6239 CDNA |
| 27 | 967.4 | 58.6 | 2974 | 25 | ABX98413 | Human CDNA encodin |
| 28 | 967.4 | 58.6 | 2974 | 25 | ABX98915 | Novel human secret |
| 29 | 967.4 | 58.6 | 2974 | 25 | ACA05960 | Human secreted/tra |
| 30 | 967.4 | 58.6 | 2974 | 25 | ABX98004 | Human PRO polynuc |
| 31 | 967.4 | 58.6 | 2974 | 25 | ABX78788 | Human PRO polynuc |
| 32 | 967.4 | 58.6 | 2974 | 25 | ABX75801 | Human CDNA encodin |
| 33 | 967.4 | 58.6 | 2974 | 25 | ABX77006 | Human PRO polynuc |
| 34 | 967.4 | 58.6 | 2974 | 25 | ABX16846 | Human CDNA encodin |
| 35 | 965.8 | 58.5 | 2996 | 25 | ABZ58826 | Human 32624 polype |
| 36 | 964.2 | 58.4 | 3050 | 25 | ABX72194 | Human NOVX polynuc |
| 37 | 956.8 | 58.0 | 1961 | 24 | ABK63496 | Rat sequence diffe |
| 38 | 936.8 | 56.8 | 1584 | 24 | ABN85827 | Human NOV13, glucu |
| 39 | 872.2 | 52.9 | 1606 | 24 | ABN85391 | Human NOV13, glucu |
| 40 | 836.8 | 50.7 | 1024 | 25 | ABZ83776 | Toxicologically re |
| 41 | 829.2 | 50.3 | 1766 | 23 | AA591380 | DNA encoding novel |
| 42 | 818 | 49.6 | 1413 | 23 | AA033928 | Human UDP-glucosyl |
| 43 | 754.8 | 45.7 | 1323 | 21 | AAZ95193 | Human UGT2B4 exon |
| 44 | 610.8 | 37.0 | 1686 | 21 | AAZ95201 | Human UGT2B7 exon |
| 45 | 572.4 | 34.7 | 1001 | 21 | AAH51512 | Human UGT2B10 rela |

ALIGNMENTS

RESULT 1
AAC65396
ID AAC65396 standard; CDNA; 1650 BP.
AC AAC65396;
XX
XX
XX 13-FEB-2001 (first entry)
XX
DE Human carbohydrate-modifying enzyme CDNA Incyte ID No: 2912330CB1.
XX
XX Human, carbohydrate-modifying enzyme; CME; antidiabetic;
XX immunosuppressive; anti-HIV; antiinflammatory; antiaiemic;
XX antiaethmatic; antiarteriosclerotic; antithyroid; hepatoprotic;
XX nephrotropic; antipour; thymimetic; neuroprotective; osteopahic;
XX antiaethmatic; antipour; thymimetic; neuroprotective; osteopahic;
XX dermatological; antidiabetic; cytosolic; vituicide; antibacterial;
XX fungicide; protozoacide; tranquilliser; vulnery; diabetes;
XX autoimmune disorder; inflammatory disorder; infection; ss.
XX
XX Homo sapiens.
XX
XX WO200063351-A2.
XX
XX 26-OCT-2000.
XX
XX 20-APR-2000; 2000WO-UG10882.
XX
XX 21-APR-1999; 99US-0130383.
XX
XX (INCY-) INCYTE GENOMICS INC.
XX

RESULT 2
 AA295199
 ID AA295199 standard; DNA; 2092 BP.
 XX
 AC AA295199;
 XX
 DT 05-JUN-2000 (first entry)
 XX
 DE Human UDP-glucuronosyltransferase 2B4 nucleotide sequence.
 XX
 KW UDP-glucuronosyltransferase 2B4; UGT2B4; polymorphism; metabolism; SNPs;
 KM drug interaction; detect; human; single nucleotide polymorphism; ds.
 OS Homo sapiens.
 XX
 FN W0200006776-A1.
 XX
 PD 10-FEB-2000.
 XX
 PF 22-JUL-1999; 99WO-US16675.
 XX
 PR 28-JUL-1998; 98US-0094391.
 XX
 PA (AXYS-) AXYS PHARM INC.
 XX
 PI Galvin M, Miller A, Penny L, Riedy M;
 XX
 DR WPI; 2000-195321/17.
 DR P-PSDB; AAY78933.
 PT Novel human UDP-glucuronosyltransferase sequence, polymorphisms for
 PT genotyping individuals to predict rate of metabolism of substrates and
 PT for identifying potential drug interactions -
 XX
 PS Disclosure; Page 34-36; 72pp; English.
 XX
 CC This sequence represents the human UDP-glucuronosyltransferase 2B4
 CC (UGT2B4) gene. UDP-glucuronosyltransferase (UGTs) are a family of
 CC enzymes that catalyze the glucuronic acid conjugation of a wide range of
 CC endogenous and exogenous substrates. The UGT2B gene subfamily encode
 CC steroid metabolizing isoforms in the liver. Alteration of the expression
 CC or function of UGTs may affect drug metabolism. The invention relates to
 CC non-chromosomal nucleic acid molecules, which comprise human UGT2B
 CC sequence polymorphisms (see AA295051-Z95110). Probes which detect the
 CC UGT2B locus polymorphisms can be used to detect altered UGT2B metabolism
 CC of a substrate in an individual. The nucleic acid molecules comprising a
 CC human UGT2B sequence polymorphism can be used in screening assays for
 CC genotyping individuals, also to predict their rate of metabolism of
 CC UGT2B substrate, potential drug-drug interactions and adverse side
 CC effects. The polymorphisms can be used as single nucleotide polymorphisms
 CC (SNPs) for detecting genetic linkage related to phenotypic variation in
 CC activity or expression of UGT2B protein. The polymorphism containing
 CC nucleic acid molecules may also be used for generating genetically
 CC modified non-human animals and for obtaining site specific gene
 CC modification in cell lines.
 XX
 SO Sequence 2092 BP; 639 A; 398 C; 438 G; 617 T; 0 other;
 Query Match 87.6%; Score 1445.4; DB 21; Length 2092;
 Best Local Similarity 93.3%; Pred. No. 0;
 Matches 1522; Conservative 0; Mismatches 106; Indels 3; Gaps 1;

| | | | |
|----|------|---|------|
| Db | 124 | GCCACAGAAATTCAGCCACTGATGAATATAAGACAATCCGTGATGATGTCACAG | 183 |
| Qy | 181 | AGGTCATGAGGTGCTGATATGGCATCTTCAGCTTCATTTCTTTCGATCCCAACAGCC | 240 |
| Db | 184 | AGGTCATGAGGTGCTGATATGGCATCTTCAGCTTCATTTCTTTCGATCCCAACAGCC | 243 |
| Qy | 241 | ATCTACTCTTAATTTGAAAGTTTATCCTGTATCTTTAACTAAAATGAGTTTGAGATAT | 300 |
| Db | 244 | ATCTACTCTTAATTTGAAAGTTTATCCTGTATCTTTAACTAAAATGAGTTTGAGATAT | 303 |
| Qy | 301 | TATCAACAGCTGTTTAAGATGCGCAGAACTTCCAAAGACACATTTTGGTCATATTT | 360 |
| Db | 304 | TATCAACAGCTGTTTAAGATGCGCAGAACTTCCAAAGACACATTTTGGTCATATTT | 363 |
| Qy | 361 | TTCCACAGTCAAGAAATCAGTGGACATTTAATGATCATCTTAAGAAAGTTCTGTAAGA | 420 |
| Db | 364 | TTCCACAGTCAAGAAATCAGTGGACATTTAATGATCATCTTAAGAAAGTTCTGTAAGA | 423 |
| Qy | 421 | TATAGTTTCAAAATTAAGAACTTATGAAGAAACTACAGAGTCAAGATTTGATGTTTCT | 480 |
| Db | 424 | TATAGTTTCAAAATTAAGAACTTATGAAGAAACTACAGAGTCAAGATTTGATGTTTCT | 483 |
| Qy | 481 | TGCAGATGCTGTTTCCCTTGGTGAGCTGCTGCGGAGTTACTTAAATACCTTTGT | 540 |
| Db | 484 | TGCAGATGCTGTTTCCCTTGGTGAGCTGCTGCGGAGTTACTTAAATACCTTTGT | 543 |
| Qy | 541 | CTACAGCTCCGCTTCTCTGCTGCTACGCAATTAAGAAAGATGATGAGAGACTTCTGT | 600 |
| Db | 544 | CTACAGCTCCGCTTCTCTGCTGCTACGCAATTAAGAAAGATGATGAGAGACTTCTGT | 603 |
| Qy | 601 | CCCTCCTTCTATGTCGCTGTTGTTATGTCAGAACTAAGTACCAATGACTTTTCATAGA | 660 |
| Db | 604 | CCCTCCTTCTATGTCGCTGTTGTTATGTCAGAACTAAGTACCAATGACTTTTCATAGA | 663 |
| Qy | 661 | GAGGTAATAAATATGATCTATGCTTTAATTTGAAATTTGGTTCCAAATATTTGACAT | 720 |
| Db | 664 | GAGGTAATAAATATGATCTATGCTTTAATTTGAAATTTGGTTCCAAATATTTGACAT | 723 |
| Qy | 721 | GAAGAAATGGATACGTTTACAGTGAAGTTCTAGAGAAAGCCCACTACGTTATCTGAGAC | 780 |
| Db | 724 | GAAGAAATGGATACGTTTACAGTGAAGTTCTAGAGAAAGCCCACTACGTTATCTGAGAC | 783 |
| Qy | 781 | AATGCAAAAGCTGACATATGCTTATTCGAAACTACTGGGATTTTCAATTTCTCTACCC | 840 |
| Db | 784 | AATGCAAAAGCTGACATATGCTTATTCGAAACTACTGGGATTTTCAATTTCTCTACCC | 843 |
| Qy | 841 | ACTCTTACCAATGTTGATGCTGTTGAGAGACTTCACCTGCAAACTGCCAAACCCCTAAC | 900 |
| Db | 844 | ACTCTTACCAATGTTGATGCTGTTGAGAGACTTCACCTGCAAACTGCCAAACCCCTAAC | 903 |
| Qy | 901 | GAAGAAATGGAAGATTTTGTCCAGAGCTCTGAGAAATGCTGTTGTTCTCT | 960 |
| Db | 904 | GAAGAAATGGAAGATTTTGTCCAGAGCTCTGAGAAATGCTGTTGTTCTCT | 963 |
| Qy | 961 | GGGTCATGCTGCTAAGACAGTCAAGAAAGGCAATGTAATTCATCAGCCCTTGC | 1020 |
| Db | 964 | GGGTCATGCTGCTAAGACAGTCAAGAAAGGCAATGTAATTCATCAGCCCTTGC | 1023 |
| Qy | 1021 | CAAGATCCACAAAGGTTCTGTGAGATTTGATGGAATTAACCAAGATCTTTAGAGACT | 1080 |
| Db | 1024 | CAAGATCCACAAAGGTTCTGTGAGATTTGATGGAATTAACCAAGATCTTTAGAGACT | 1083 |
| Qy | 1081 | CAATATCTGCTGCTAAGACAGTCAAGAAAGGCAATGTAATTCATCAGCCCTTGC | 1140 |
| Db | 1084 | CAATATCTGCTGCTAAGACAGTCAAGAAAGGCAATGTAATTCATCAGCCCTTGC | 1143 |
| Qy | 1141 | AGCTTTATCACTATGTTGGAATGATGGAATCTATGAGCTATTTACCATGGGCTCC | 1200 |
| Db | 1144 | AGCTTTATCACTATGTTGGAATGATGGAATCTATGAGCTATTTACCATGGGCTCC | 1203 |
| Qy | 1201 | TATGTTGGAGTTCCCATATTTTGGTATCAGCTTGATTAACATGACTCAATGAAGGCCAA | 1260 |
| Db | 1204 | TATGTTGGAGTTCCCATATTTTGGTATCAGCTTGATTAACATGACTCAATGAAGGCCAA | 1263 |


```
Db 844 ACTTTACCAATGTTGAGTTCGTTGGAGAGACTCCACATGCAAACTGCGCAAACTCCCTACC 903
Qy 901 GAAGGAATGGAAGAGTTTGTCCAGAGCTCGAGAGAAAATGGTGTGTTGTTCTCT 960
Db 904 GAAAGAAATGGAAGAGTTTGTCCAGAGCTCGAGAGAAAATGGTGTGTTGTTCTCT 963
Qy 961 GGGGTGATGTGTCAGTAACAGTCAGAGAAAGAGCCAAATGTAATGGATCAGCCCTTGC 1020
Db 964 GGGGTGATGTGTCAGTAACAGTCAGAGAAAGAGCCAAATGTAATGGATCAGCCCTTGC 1023
Qy 1021 CAAGATCCCAAAAAGTTCTGTGAGATTTGATGGGAATTAACCAAGATCTTATGAGACT 1080
Db 1024 CAAGATCCCAAAAAGTTCTGTGAGATTTGATGGGAATTAACCAAGATCTTATGAGACT 1083
Qy 1081 CAATACCTGGGCTGTCAGAGTGAATCCCAAGATATCTTCTGGTCAATCCCAAAACCA 1140
Db 1084 CAATACCTGGGCTGTCAGAGTGAATCCCAAGATATCTTCTGGTCAATCCCAAAACCA 1143
Qy 1141 AGCTTTTATCACTCATGTGGAATGATGATGATGATGATGATGATGATGATGATGATGAT 1200
Db 1144 AGCTTTTATCACTCATGTGGAATGATGATGATGATGATGATGATGATGATGATGATGAT 1203
Qy 1201 TATGTGGGAGTTCCTCATATTTTGTGATCACTGATTAACATAGCTCAATGAGGCCAA 1260
Db 1204 TATGTGGGAGTTCCTCATATTTTGTGATCACTGATTAACATAGCTCAATGAGGCCAA 1263
Qy 1261 AGGAGCAGCTGTAGAAATTAACCTTAACTTAATGATGATGATGATGATGATGATGATGAT 1320
Db 1264 GGGAGCAGCTGTAGAAATTAACCTTAACTTAATGATGATGATGATGATGATGATGATGAT 1323
Qy 1321 GAGAAGCAGCTGTAGAAATTAACCTTAACTTAATGATGATGATGATGATGATGATGATGAT 1380
Db 1324 GAGAAGCAGCTGTAGAAATTAACCTTAACTTAATGATGATGATGATGATGATGATGATGAT 1383
Qy 1381 CCATGATCAACCTGTAAGAGCCCTTGTAGATGAGAGCTCTTGTGATGAGATTTGTCATGCG 1440
Db 1384 TCATGATCAACCTGTAAGAGCCCTTGTAGATGAGAGCTCTTGTGATGAGATTTGTCATGCG 1443
Qy 1441 CCACAAAGAGCCCAAGAGCTGCGATGAGCTGCGCATACCTCACTGCTGTCAGACATA 1500
Db 1444 CCATTAAGAGCCCAAGAGCTGCGATGAGCTGCGCATACCTCACTGCTGTCAGACATA 1503
Qy 1501 CTCTATAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1560
Db 1504 CTCTATAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1563
Qy 1561 AAAATGTTTATTTTCTGTCAAAATTTAATTAATTAATTAATTAATTAATTAATTAATTA 1620
Db 1564 AAAATGTTTATTTTCTGTCAAAATTTAATTAATTAATTAATTAATTAATTAATTAATTA 1620
Qy 1621 ATAGATCTTTC 1631
Db 1621 TTAATTTAGCT 1631

RESULT 4
ID AAS69712 standard; cDNA; 2111 BP.
AC AAS69712;
XX
XX 13-FEB-2002 (first entry)
XX
XX DNA encoding novel human diagnostic protein #5516.
XX
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
XX Homo sapiens.
XX
XX WO200175067-A2.
XX
```

```
PD 11-OCT-2001.
XX
XX 30-MAR-2001; 2001WO-US08631.
XX
XX 31-MAR-2000; 2000US-0540217.
XX
XX 23-AUG-2000; 2000US-0649167.
XX
XX (HYSE-) HYSEQ INC.
XX
XX Dmanac RT, Liu C, Tang YT;
XX
XX WPI; 2001-639362/73.
XX
XX P-PSDB; ABG05525.
XX
XX New isolated polynucleotide and encoded polypeptides, useful in
XX diagnostics, forensic, gene mapping, identification of mutations
XX responsible for genetic disorders or other traits and to assess
XX biodiversity
XX
XX Claim 1; SEQ ID No 5516; 103pp; English.
XX
XX The invention relates to isolated polynucleotide (I) and
XX polypeptide (II) sequences. (I) is useful as hybridisation probes,
XX polymerase chain reaction (PCR) primers, oligomers, and for chromosome
XX and gene mapping, and in recombinant production of (II). The
XX polynucleotides are also used in diagnostics as expressed sequence tags
XX for identifying expressed genes. (I) is useful in gene therapy techniques
XX to restore normal activity of (II) or to treat disease states involving
XX (II). (II) is useful for generating antibodies against it, detecting or
XX quantitating a polypeptide in tissue, as molecular weight markers and as
XX a food supplement. (II) and its binding partners are useful in medical
XX imaging of sites expressing (II). (I) and (II) are useful for treating
XX disorders involving aberrant protein expression or biological activity.
XX The polypeptide and polynucleotide sequences have applications in
XX diagnostics, forensic, gene mapping, identification of mutations
XX responsible for genetic disorders or other traits to assess biodiversity
XX and to produce other types of data and products dependent on DNA and
XX amino acid sequences. AAS64197-AAS94564 represent novel human
XX diagnostic coding sequences of the invention.
XX Note: The sequence data for this patent did not appear in the printed
XX at ftp.wipo.int/pub/published_pcl_sequences.
XX
XX Sequence 2111 BP; 655 A; 397 C; 442 G; 617 T; 0 other;
XX
XX Query Match 86.5%; Score 1428; DB 23; Length 2111;
XX Best Local Similarity 93.0%; Pred. No. 0;
XX Matches 1518; Conservative 0; Mismatches 110; Indels 4; Gaps 2;
XX
Qy 1 AGCACTGGAAGAAACAAGCATTTGATGATGATGATGATGATGATGATGATGATGATGATGAT 60
Db 4 AGCACTGGAAGAAACAAGCATTTGATGATGATGATGATGATGATGATGATGATGATGATGAT 63
Qy 61 TCTCTGATACAGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 120
Db 64 TCTCTGATACAGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 123
Qy 121 GCCCAGAAATTCAGCCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 180
Db 124 GCCCAGAAATTCAGCCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 183
Qy 181 AGGTCATGAGTGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 240
Db 184 AGGTCATGAGTGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 243
Qy 241 ATCTACTCTTAATTTGAATTTATCTGATCTTTTAACTTAATTAATTAATTAATTAATTAAT 300
Db 244 ATCTACTCTTAATTTGAATTTATCTGATCTTTTAACTTAATTAATTAATTAATTAATTAAT 303
Qy 301 TATCAGCAGCTGTTAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 360
Db 304 TATCAGCAGCTGTTAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 363
```

```

Qy 361 TTCAAGATACAGAAATCATGTGAGCATTTAATGACATCTAGAAAAGTTCTGTAAAGA 420
Db 364 TTCAAGATACAGAAATCATGTGAGCATTTAATGACATCTAGAAAAGTTCTGTAAAGA 423
Qy 421 TATAGTTTCAATAAAGAACTTATGAGAAAATCAAGAGTCAGATTTGATGTTTCT 480
Db 424 TATAGTTTCAATAAAGAACTTATGAGAAAATCAAGAGTCAGATTTGATGTTTCT 483
Qy 481 TGCAGATCTGTTTTCCCTTTGAGAGCTGTGGCGAGTTACTTAAATACCTTTGT 540
Db 484 TGCAGATCTGTTTTCCCTTTGAGAGCTGTGGCGAGTTACTTAAATACCTTTGT 543
Qy 541 CTACAGCTCCGCTTCTCTCTGAGTACCAATTGAAAGCAATGTGAGAGACTTCTGT 600
Db 544 CTACAGCTCCGCTTCTCTCTGAGTACCAATTGAAAGCAATGTGAGAGACTTCTGT 603
Qy 601 CCTCTCTTCTATGTCCTGTTGTTATGTAGAACTAAGTACCAATGACTTTCAATGA 660
Db 604 CCTCTCTTCTATGTCCTGTTGTTATGTAGAACTAAGTACCAATGACTTTCAATGA 663
Qy 661 GAGGTTAAAAAATAGATCTATGCTTATTTGAAATTTGGTCCAAATATTTGACAT 720
Db 664 GAGGTTAAAAAATAGATCTATGCTTATTTGAAATTTGGTCCAAATATTTGACAT 723
Qy 721 GAAGAAGTGGATCAGTCTACAGTGAAGTTCTAGAAAGACCCACTAGCTTATCTGAGAC 780
Db 724 GAAGAAGTGGATCAGTCTACAGTGAAGTTCTAGAAAGACCCACTAGCTTATCTGAGAC 783
Qy 781 AATGGCAAAAGCTGACATATGCTTATTTGAAACTACTGGGATTTTCAATTTCTCACC 840
Db 784 AATGGCAAAAGCTGACATATGCTTATTTGAAACTACTGGGATTTTCAATTTCTCACC 843
Qy 841 ACTCTTACCAATGTTGAGTTCGTGGAGAACTCCAGCTGCAAACTGGCAAAACCCCTAC 900
Db 844 ACTCTTACCAATGTTGAGTTCGTGGAGAACTCCAGCTGCAAACTGGCAAAACCCCTAC 903
Qy 901 GAAGAAGTGAAGTGAAGTTCGTGGAGAACTCCAGCTGCAAACTGGCAAAACCCCTAC 959
Db 904 GAAGAAGTGAAGTGAAGTTCGTGGAGAACTCCAGCTGCAAACTGGCAAAACCCCTAC 963
Qy 960 TGGGTCGATGTCAGTAAACAGTCAGAGAAAGGCGCAATGTAATTCATCAGCCCTTG 1019
Db 964 TGGGTCGATGTCAGTAAACAGTCAGAGAAAGGCGCAATGTAATTCATCAGCCCTTG 1023
Qy 1020 CCAAGATCCCAAAAGGTTCTGTGAGATTTGATGGGAAATTAACCAATCTTTAGAC 1079
Db 1024 CCAAGATCCCAAAAGGTTCTGTGAGATTTGATGGGAAATTAACCAATCTTTAGAC 1083
Qy 1080 TCAATACTCGCTGTACAAAGTGGATACCCAGATGATCTTGTGTCATCCCAAAACCA 1139
Db 1084 TCAATACTCGCTGTACAAAGTGGATACCCAGATGATCTTGTGTCATCCCAAAACCA 1143
Qy 1140 AAGCTTTTATCACTCATGTTGGAATGAATGGGATCTATGAGCTTATTAACATGAGG 1199
Db 1144 AAGCTTTTATCACTCATGTTGGAATGAATGGGATCTATGAGCTTATTAACATGAGG 1203
Qy 1200 CTATGTTGGAGATTTCCCAATTTTGTGATGACCTTGAATCACTCAATGAGAGGCA 1259
Db 1204 CTATGTTGGAGATTTCCCAATTTTGTGATGACCTTGAATCACTCAATGAGAGGCA 1263
Qy 1260 AAGGAGCAGCTGTAGAAATTAATCTCAAAATCATGCAAGCAAGATTTACTGAGGCTT 1319
Db 1264 AAGGAGCAGCTGTAGAAATTAATCTCAAAATCATGCAAGCAAGATTTACTGAGGCTT 1323
Qy 1320 TGAGAACAGTCACTTACCGATTCCTCTTATTAAGAGATCTATGAGATTTATCAAGATTC 1379
Db 1324 TGAGAACAGTCACTTATGATCTTTATATTAAGAGATCTATGAGATTTATCAAGATTC 1383
Qy 1380 ACCATGATTAACCTGTAAAGCCCCCTTATGATGAGCACTCTTCTGATCGAGTTTTCATCC 1439
Db 1384 ACCATGATTAACCTGTAAAGCCCCCTTATGATGAGCACTCTTCTGATCGAGTTTTCATCC 1443
Qy 1440 GCCAAGAAAGGAGCAAGCACTGTGATGATGATGATGATGATGATGATGATGATGATG 1499

```

```

Db 1444 GCCAAGAAAGGAGCAAGCACTGTGATGATGATGATGATGATGATGATGATGATG 1503
Qy 1500 ACTCTATAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1559
Db 1504 ACTCTATAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1563
Qy 1560 CAAATGTTTTTATTTTCTCTGTCMAAAATTTAATAAAGTATGAAAGTAAAGAGG 1619
Db 1564 CAAATGTTTTTATTTTCTCTGTCMAAAATTTAATAAAGTATGAAAGTAAAGAGG 1620
Qy 1620 AATGATCTTTC 1631
Db 1621 AATGATCTTTC 1632

RESULT 5
AAL41490
ID AAL41490 standard; DNA; 1639 BP.
XX
AC AAL41490;
XX
DT 21-NOV-2002 (first entry)
XX
DE Drug metabolizing enzyme encoding DNA - 7493833CB1.
XX
KW Anti-HIV; antiarteriosclerotic; dermatological; cytosolic; thymimetic;
KW osteopathic; thrombolytic; ophthalmological; antileptic; hepatocarcinoma;
KW drug metabolizing enzyme; DME; inflammatory; AIDS; atherosclerosis;
KW contact dermatitis; cell-proliferative; cancer; cirrhosis; dwarfism;
KW developmental; hypothyroidism; endocrine; osteoporosis; thrombosis;
KW diabetes; glaucoma; keratitis; metabolic; hyperlipidemia; diarrhoea;
KW cystic fibrosis; gastroenteritis; gastroenteritis; liver; hepatitis;
KW Reye's syndrome; exogenous compound; gene therapy; enzyme; human; gene;
KW ds.
XX
XX Homo sapiens.
XX
OS
XX
FH Key 1..1590 Location/Qualifiers
FT CDS /*tag= a /product= "Drug metabolizing enzyme protein"
FT
FT
XX
XX MO20026654-A2.
XX
XX 29-AUG-2002.
XX
XX 14-FEB-2002; 2002MO-US04918.
XX
XX 16-FEB-2001; 2001US-269643P.
XX
XX 23-FEB-2001; 2001US-271332P.
XX
XX 16-MAR-2001; 2001US-276767P.
XX
XX 06-APR-2001; 2001US-282077P.
XX
XX 19-APR-2001; 2001US-285447P.
XX
XX 27-APR-2001; 2001US-287060P.
XX
XX 03-MAY-2001; 2001US-288543P.
XX
XX (INCY-) INCYTE GENOMICS INC.
XX
XX Astromoff A, Au-Yang J, Baughn MR, Ding L, Duggan BM;
XX Forrethe J, Gietzen KJ, Griffen JA, Lee EA, Lu Y, Richardson TW;
XX Ring HZ, Sanjanwala MM, Swarnakar A, Walla NK, Warren BA, Xu Y,
XX Yue H, Zebardjian Y;
XX
XX WPI; 2002-674949/72.
XX
XX P-PeDB; AAO22794.
XX
XX New drug metabolizing enzymes (DME) useful for diagnosing, treating and
XX preventing diseases or conditions associated with aberrant DME
XX expression, e.g. cancer, AIDS, atherosclerosis, diabetes, glaucoma,
XX hepatitis, osteoporosis -
XX

```

PS Claim 5; Page 161-162; 166bp; English.

XX The invention relates to an isolated polypeptide comprising 12 fully defined sequences of 81-615 amino acids given in the specification, a naturally occurring amino acid sequence at least 90% identical to, having 81-599 amino acids, at least 93% identical to a sequence of 529 amino acids, or at least 97% identical to a sequence of 615 amino acids, all given in the specification; or a biologically active or immunogenic fragment of the polypeptide. The polypeptides and polynucleotides are useful in diagnosing, treating and preventing diseases or conditions associated with the decreased expression or overexpression of a drug metabolizing enzyme (DME), such as autoimmune/inflammatory (e.g. AIDS, attherosclerosis, contact dermatitis) cell-proliferative (e.g. cancer, cirrhosis), developmental (e.g. diabetes, hypochyroidism), endocrine (e.g. osteoporosis, thrombosis, eye (e.g. glaucoma, keratitis), metabolic (e.g. hyperlipidaemia, cystic fibrosis), galectinosis (e.g. gastroenteritis, diarrhoea), or liver (e.g. hepatitis, Reye's syndrome) disorders. These are also useful in assessing the effects of exogenous compounds on the expression of nucleic acid and amino acid sequences of DME. The DME or its fragments are useful in screening compounds for effectiveness as agonist or antagonist of the polypeptides, or in altering the expression of the target polynucleotide and compounds that specifically bind to or modulate the activity of the polypeptide. The microarray is useful in monitoring or measuring protein-protein interactions, drug-target interactions, and gene expression profiles. This polynucleotide sequence represents the DNA encoding a drug metabolizing enzyme of the invention.

CC Sequence 1639 BP; 470 A; 316 C; 372 G; 481 T; 0 other;

XX Query Match 80.7%; Score 1331.4; DB 24; Length 1639;
Best Local Similarity 89.6%; Pred. No. 0;
Matches 1431; Conservative 0; Mismatches 166; Indels 0; Gaps 0;

QY 35 ATGCTATGAATGAGACTTCACTTCTCTGATACAGCTGATGTTACTTAACTCT 94
Db 1 ATGCTATGAATGAGACTTCACTTCTCTGATACAGCTGATGTTACTTAACTCT 60

QY 95 GGGAGTTGTGAAAAGTCTGTGTGGCCCAAGAAATTCAGCCACTGATGAATATAAG 154
Db 61 GGGAGTTGTGAAAAGTCTGTGTGGCCCAAGAAATTCAGCCACTGATGAATATAAG 120

QY 155 ACAATCTCGATGAGACTTGTCCAGAGAGATGATGAGTATGGATCTTCACT 214
Db 121 ACAATCTCGATGAGACTTGTCCAGAGAGATGATGAGTATGGATCTTCACT 180

QY 215 TCCATTTCTTTCGATCCCAAGCCCATCTACTTTAAATTTGAATTTATCTGATCT 274
Db 181 TCCATTTCTTTCGATCCCAAGCCCATCTACTTTAAATTTGAATTTATCTGATCT 240

QY 275 TTAATCTAAACTGATTTGAGATATTAACAAGAGCTGTTAAAGATGGCGAACTT 334
Db 241 TTAATCTAAACTGATTTGAGATATTAACAAGAGCTGTTAAAGATGGCGAACTT 300

QY 335 CCAAAAGACATTTTGGTCAATTTTCAAGTACAGAAATCATCTGACATTTAAT 394
Db 301 CCAAAAGACATTTTGGTCAATTTTCAAGTACAGAAATCATCTGACATTTAAT 360

QY 395 GACATCTTAGAAAGTTCTGTAGAGATATAGTTTCAATATAGAACTTAAGAAAATA 454
Db 361 GACATCTTAGAAAGTTCTGTAGAGATATAGTTTCAATATAGAACTTAAGAAAATA 420

QY 455 CAGAGATCAAGATTTGATGTTGTTCTTGACAGATGCTGTTTCCCTTTGGAGCTGTG 514
Db 421 CAGAGATCAAGATTTGATGTTGTTCTTGACAGATGCTGTTTCCCTTTGGAGCTGTG 480

QY 515 GCCAGTACTTAATAATCCCTTGTCTACAGCCCTCGCTCTCTCCGAGCTAGCAATT 574
Db 481 GCTGAGCTATTTAAATACCTTTGTGTACAGCTCAGCTTCTCCGAGCTAGCAATT 540

QY 575 GAAAGCATAGTGAGAGACTTCTGTTCCCTCTCTATGCTGCTGTATGTACAGAA 634
Db 541 GAAAGCATAGTGAGAGACTTCTGTTCCCTCTCTATGCTGCTGTATGTACAGAA 600

Db 541 GAAAGCATAGTGAGAGACTTCTGTTCCCTCTCTATGCTGCTGTATGTACAGAA 600

QY 635 CTATGACCAAAATGACTTTCATAGAGGGTAAATAATGATCTATGCTTATTTT 694
Db 601 TTAATCTATCAAAATGACTTTCATAGAGGGTAAATAATGATCTATGCTTATTTT 660

QY 695 GAATTTGTTTCCAAATATTTTGAATGAAGAGTGGATGATCTTACAGTAAATTCTA 754
Db 661 GACTTTGTTTCCAAATATTTTGAATGAAGAGTGGATGATCTTATGTAAGATTCTA 720

QY 755 GGAAGACCACTACGTTATCTGAGACATGCGAAAGCTGACATATGCTTATTCGAAC 814
Db 721 GGAAGACCACTACGTTATCTGAGACATGCGAAAGCTGACATATGCTTATTCGAAC 780

QY 815 TACGGAATTTTCAATTTTCTCCACCCCTTACCAATGTTGAGTTCGTGAGGACTC 874
Db 781 TCCGGAATTTTCAATTTTCTCCACCCCTTACCAATGTTGAGTTCGTGAGGACTC 840

QY 875 CACTGCAAACTGCGCAAAACCCCTACCGAAGAAATGGAAGTTTTCAGAGCTCTGGA 934
Db 841 CACTGCAAACTGCGCAAAACCCCTACCGAAGAAATGGAAGTTTTCAGAGCTCTGGA 900

QY 935 GAAATGTTGTTGTTGTTTCTGCGGATCGATGTCAGTAAACAGTCAGAAAGG 994
Db 901 GAAATGTTGTTGTTGTTTCTGCGGATCGATGTCAGTAAACAGTCAGAAAGG 960

QY 995 GCCAATATATTTGATAGCCCTTGGCCAAAGATCCCAAAAGTTCTGTGGAATTTGAT 1054
Db 961 GCCAATATATTTGATAGCCCTTGGCCAAAGATCCCAAAAGTTCTGTGGAATTTGAT 1020

QY 1055 GGGATTAACAGATTAATTTAGAGACTCAATATCTGCGCTGTACAAAGTGGATACCCAGAA 1114
Db 1021 GGGATTAACAGATTAATTTAGAGACTCAATATCTGCGCTGTACAAAGTGGATACCCAGAA 1080

QY 1115 GATCTTCTTGTGATCCCAAAACCAAGCTTTTATCACTGATGTTGAATGATGGATC 1174
Db 1081 GATCTTCTTGTGATCCCAAAACCAAGCTTTTATCACTGATGTTGAATGATGGATC 1140

QY 1175 TATGAGCTATTTTACATGAGGATCCCTATATGTTGGAGTTCCCATATTTGGTATCAGCTT 1234
Db 1141 TATGAGCTATTTTACATGAGGATCCCTATATGTTGGAGTTCCCATATTTGGTATCAGCTT 1200

QY 1235 GATTAACATGCTCAATGAAGAGCCCAAGAGAGAGCTGTAGAATTAATCAAAACTATG 1294
Db 1201 GATTAACATGCTCAATGAAGAGCCCAAGAGAGAGCTGTAGAATTAATCAAAACTATG 1260

QY 1295 ACAAGCGAATTTTACTGAGGGCTTTGAGAACAGTCAATTCAGATTCCTTTTAAAGAG 1354
Db 1261 TCGAGTACAGATTTTCTGAAATGATTAAGAGATTAATGATTCCTTTTAAAGAG 1320

QY 1355 AATGCTATGAGATTTTCAAGAAATTTACCATGATCAACCTGTAAGCCCTAGATCGAGCA 1414
Db 1321 AATGCTATGAGATTTTCAAGAAATTTCAACATGATCAACCATGTAAGCCCTGATCGAGCA 1380

QY 1415 GTCTTCGATGAGTTTGTGATGCGCCCAAGAGAGCCCAAGACCTGCGATCAGCTGCC 1474
Db 1381 GTCTTCGATGAGTTTGTGATGCGCCCAAGAGAGCCCAAGACCTGCGATCAGCTGCC 1440

QY 1475 CATGACCTCACTGTTCCAGACACTACTTATAGATGTAATGAGTTCTGCTGACCTGT 1534
Db 1441 CAGAGCTCACTGTTCCAGACACTCTTTGATGTAATGAGTTCTGCTGCTGTG 1500

QY 1535 GTGGCAACTGCTATATTTCTTGTTCACAAATGTTTATTTTCTGCTGCAAAATTTAAT 1594
Db 1501 GTGGCAACTGCTATATTTCTTGTTCACAAATGTTTCTGCTGCTGCAAAATTTGTT 1560

QY 1595 AAAACTAGAAAGATAGAAAGAGGGAATAGATCTTTC 1631
Db 1561 AGAATAGAAAGAGGGAATAGATTAATTAATGCTC 1597

RESULT 6
AAD45991

QY 1381 CCATGATCAACCTGTAAAGCCCTAGATCGAGAGTCTTCTGATCGAGTTTGTCTATGCG 1440
 DB 1497 ACATGATCAACCATGGAAGCCCTGATCGAGAGTCTTCTGATGATTTGTCATGCG 1556
 QY 1441 CCACAAAGAGCCCAAGACCTGCGATCAGCTGCCATGACCTCGTTCCAGCACTA 1500
 DB 1557 CCACAAAGAGCTTAAACACCTTCGGGTTCGAGCCCAAGCACTCCTGTTCCAGTACA 1616
 QY 1501 CTCATGATGATGATGGGTTCCTGCTGATCTGTGTGGCACTGCTATTTCTTGTTCAC 1560
 DB 1617 CTTTGGATGATGATGGGTTCCTGCTGATCTGTGTGGCACTGCTATTTATTCGTCAC 1676
 QY 1561 AAAATGTTTTTATTTTCTGTCGCAAAAATTTAATAAACTGAAGAATAGAAAGAGGGA 1620
 DB 1677 AAAATGTTGCTGTTTGTCTTCTGGAAGTTTGTCTAGAAACCAAGAGGAAAAATGA 1736
 QY 1621 ATGATCTTTTC 1631
 DB 1737 TTAGTTATATC 1747

RESULT 7

ABN95622
 ID ABN95622 standard; DNA; 1855 BP.

XX ABN95622;

DT 13-AUG-2002 (first entry)

DE Gene #2120 used to diagnose liver cancer.

KW Gene: liver cancer; ds; hepatocellular carcinoma; hepatotropic;
 KM metastatic liver tumour; cytostatic; expression profile; disease state;
 XX disease progression; drug toxicity; drug efficacy; drug metabolism.

OS Homo sapiens.

PN W0200229103-A2.

XX 11-APR-2002.

PF 02-OCT-2001; 2001WO-US30589.

PR 02-OCT-2000; 2000US-237054P.

XX (GENE-) GENE LOGIC INC.

PI Horne D, Alvares C, Peres-Da-Silva S, Vockley JG;

DR WPI; 2002-426119/45.

PT Diagnosing and detecting the progression of liver cancer,
 PT hepatocellular carcinoma or metastatic liver tumor in a patient,
 PT involves detecting the level of expression of two or more genes in a
 PT liver tissue sample -

PS Claim 1; SEQ ID NO 2120; 298bp; English.

XX The invention relates to a novel method for diagnosing and detecting the
 CC progression of liver cancer, hepatocellular carcinoma or metastatic liver
 CC tumour in a patient, and differentiating metastatic liver cancer from
 CC hepatocellular carcinoma in a patient, involving detecting the level of
 CC expression of two or more genes represented in ABN93503-ABN97455 in a
 CC tissue sample. The method of the invention has hepatotropic, and
 CC cytosolic activity. The method is useful for diagnosing and detecting
 CC the progression of liver cancer, hepatocellular carcinoma and metastatic
 CC liver carcinoma in a patient. The method is useful for identifying
 CC expression profiles which serve as useful diagnostic markers as well as
 CC markers that can be used to monitor disease states, disease progression,
 CC drug toxicity, drug efficacy and drug metabolism.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO

CC at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 1855 BP; 571 A; 337 C; 394 G; 553 T; 0 other;

SO Query Match 75.3%; Score 1252.6; DB 24; Length 1855;

Best Local Similarity 86.1%; Pred. No. 0;

Matches 1387; Conservative 0; Mismatches 224; Indels 0; Gaps 0;

QY 21 TGCATTCATCAGAGATGCTATGAAATGACCTTCAGTCTTCCGATACAGTGAGCT 80
 DB 1 TGCATTCACACAGAGATGCTATGAAATGACCTTCAGTATTTGCTATACACTGAGCT 60
 QY 81 GTTACTTTAGCTCTGGAGAGTTGTGAAAGGTGTGTGTGCGCCACAGAAATTCAGCCACT 140
 DB 61 TTTGCTTTACTCTCGGAAATTTGTGAAAGGTGTGTGTGTGCGGACAGAGATTCAGCCATT 120
 QY 141 GGATGAATATTAAGACATCTGTGATGAATCTTCCAGAGAGTTCATGAGTGACTGTAT 200
 DB 121 GGATGAATATTAAGACATCTGTGATGAATCTTATTCAGAGAGTTCATGAGTGACTGTAT 180
 QY 201 TGGCATCTCAGCTTCATTTCTTTGATCCCAAGCCCATCTACTATTAATTTGAAG 260
 DB 181 TGGCATCTCAGCTTCATTTCTTTGATCCCAAGCATCTATCCGCTCTTAAATTTGAAA 240
 QY 261 TTTATCCTGATCTTTAACTAAATGAGTTTGAAGATATTAACAAGCACTGTGTTAAGA 320
 DB 241 TTTATCCACATCTTTAACTAAATGAGTTTGAAGATATTAACAAGCACTGTGTTAAGA 300
 QY 321 GATGGCAGAACTTCCAAAAGACATTTTGTGATATTTTTCACAAAGTACAAAGAAATCA 380
 DB 301 GATGGTCAGACTTCCAAAAGATACATTTTGTGATATTTTTCACAAAGTACAGAAATCA 360
 QY 381 TGTGGAATTTAATGACATCTTAAGAAAGTCTGTGAAGATATAGTTTCAATTAAGAAC 440
 DB 361 TGTCAATATTTGTGACATACATACTGAAGAGTTCTGAAGATATAGTTTCAATTAAGAAC 420
 QY 441 TTATGAAGAAACTACAGAGAGTCAAGATTTGATGTTGTTCTTGAGAGTGTGTTTCCCT 500
 DB 421 TTATGAAGAAAGTACAGAGAGTCAAGATTTGATGTTGTTCTTGAGAGTGTGTTTCCCT 480
 QY 501 TTGATGAGCTGTGCGCAGTTACTTAATAATACCTTTGTCTACAGCTTCGCTCTCTC 560
 DB 481 GTAGTGAGCTGTGCGCAGTTACTTAATAATACCTTTGTCTACAGCTTCGCTCTCTC 540
 QY 561 CTGCTACGCAATTTGAAAGAGTACAGAGAGTCTGTTCCCTCCCTCTTAAGTGTCTG 620
 DB 541 CTGCTACGCAATTTGAAAGAGTACAGAGAGTCTGTTCCCTCCCTCTTAAGTGTCTG 600
 QY 621 TTGTTATGTCAGAACTACAGAGAGTCAAGTCTTCTCATAGAGAGGTAAATAATATGATCT 680
 DB 601 TTGTTATGTCAGAACTACAGAGAGTCAAGTCTTCTCATAGAGAGGTAAATAATATGATCT 660
 QY 681 ATGTGCTTTATTTTGAATTTTGTCCAAATATTTGACATGAAGAGTGGATCATGTTCT 740
 DB 661 ATGTGCTTTACTTGAATTTTGTCCAAATATTTGACATGAAGAGTGGATCATGTTCT 720
 QY 741 ACAAGTAAGTCTACAGAGAGCCCACTACCTTATCTGAGAGAGTGAAGAGTGCAGCAT 800
 DB 721 ATAGTGAAGTCTACAGAGAGCCCACTACCTTATCTGAGAGAGTGAAGAGTGCAGCAT 780
 QY 801 GGCTTATTTCAAACTACTGGAGATTTTCAATTTCTCCACCACTTTTCAAAATGTTAGT 860
 DB 781 GGCTTATTTCAAACTACTGGAGATTTTCAATTTCTCCACCACTTTTCAAAATGTTAGT 840
 QY 861 TCGTTGAGAGCTTCACCTGCAAACTGCAAACTGCAAACTGCAAACTGCAAACTGCAAACT 920
 DB 841 TTTGTTGAGAGACTCCACTGCAAACTGCAAACTGCAAACTGCAAACTGCAAACTGCAAACT 900
 QY 921 TCCAGAGCTTGGAGAGAAATGAGTGTGTGTGTTTCTCTGAGGTCGATGTCAGTACA 980
 DB 901 TACAGAGCTTGGAGAGAAATGAGTGTGTGTGTTTCTCTGAGGTCGATGTCAGTACA 960
 QY 981 CGTCAGAGAGAAAGGCCAATGTAATTCATCAGCCCTTGCAGAGATCCCAAGAGGTTTC 1040

```

Db      961  TGAAGAGAAAGGCGCACTATTTGATAGGCGCTGGCCCATGCCAATAACAAAGGTTG 1020
Qy      1041  TGTGAGATTTGATGGAGATTAACCAAGATCTTTAGACTCAATACCTGGCTGTACAAAGT 1100
Db      1021  TGTGAGATTTGATGGAGATTAACCAAGATCTTTAGACTCAATACCTGGCTGTACAAAGT 1080
Qy      1101  GGAATACCCCAAGATCTTTGATAGCTTCCCAAGCAAGCTTTTATCACTCATAGGTG 1160
Db      1081  GGATACCCCAAGATCTTTGATAGCTTCCCAAGCAAGCTTTTATCACTCATAGGTG 1140
Qy      1161  GAATGAATGGATCTATGAGCTATTTACATGGGGTCCCTATGATGGAGTCCCATAT 1220
Db      1141  GAGCAATGGATCTATGAGCTATTTACATGGGGTCCCTATGATGGAGTCCCATATG 1200
Qy      1221  TTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1280
Db      1201  TTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1260
Qy      1281  ACTTCAAACTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1340
Db      1261  ACTTCAAACTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1320
Qy      1341  CCTCTTAAAGAGATCTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1400
Db      1321  CCTCTTAAAGAGATCTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1380
Qy      1401  CCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1460
Db      1381  CCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1440
Qy      1461  TGGCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1520
Db      1441  TGGCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1500
Qy      1521  TCGTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1580
Db      1501  TCGTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1560
Qy      1561  GTCAAAATTTATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1611
Db      1561  TCTGGAAGTTGCTGATGAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 1611

RESULT 8
ABL68501
ID      ABL68501 standard; DNA; 1855 BP.
XX
XX      ABL68501;
XX
XX      15-MAY-2002 (first entry)
XX
XX      Kidney cancer related gene sequence SEQ ID NO:6838.
XX
XX      Human; cancer; colon; breast; ovary; oesophagus; kidney; thyroid;
XX      stomach; lung; prostate; pancreas; carcinoma; antitumour; cancerous;
XX      cytotoxic; gene therapy; antineoplastic; Wilms tumour; adenocarcinoma;
XX      gene; ds.
XX
XX      Homo sapiens.
XX
XX      WO200194629-A2.
XX
XX      13-DEC-2001.
XX
XX      30-MAY-2001; 2001WO-US10838.
XX
XX      05-JUN-2000; 2000US-209473P.
XX      05-JUN-2000; 2000US-209531P.
XX      18-SEP-2000; 2000US-23133P.
XX      18-SEP-2000; 2000US-231617P.
XX      20-SEP-2000; 2000US-234009P.
XX      20-SEP-2000; 2000US-234034P.

```

```

PR      20-SEP-2000; 2000US-234052P.
PR      22-SEP-2000; 2000US-234509P.
PR      22-SEP-2000; 2000US-234567P.
PR      25-SEP-2000; 2000US-234923P.
PR      25-SEP-2000; 2000US-234924P.
PR      25-SEP-2000; 2000US-235077P.
PR      25-SEP-2000; 2000US-235082P.
PR      25-SEP-2000; 2000US-235134P.
PR      25-SEP-2000; 2000US-235280P.
PR      26-SEP-2000; 2000US-235637P.
PR      26-SEP-2000; 2000US-235638P.
PR      27-SEP-2000; 2000US-235711P.
PR      27-SEP-2000; 2000US-235720P.
PR      27-SEP-2000; 2000US-235840P.
PR      27-SEP-2000; 2000US-235863P.
PR      28-SEP-2000; 2000US-236028P.
PR      28-SEP-2000; 2000US-236032P.
PR      28-SEP-2000; 2000US-236033P.
PR      28-SEP-2000; 2000US-236034P.
PR      28-SEP-2000; 2000US-236109P.
PR      28-SEP-2000; 2000US-236111P.
PR      29-SEP-2000; 2000US-236842P.
PR      29-SEP-2000; 2000US-236891P.
PR      29-SEP-2000; 2000US-237172P.
PR      02-OCT-2000; 2000US-237173P.
PR      02-OCT-2000; 2000US-237278P.
PR      02-OCT-2000; 2000US-237294P.
PR      02-OCT-2000; 2000US-237295P.
PR      02-OCT-2000; 2000US-237316P.
PR      03-OCT-2000; 2000US-237425P.
PR      03-OCT-2000; 2000US-237598P.
PR      03-OCT-2000; 2000US-237604P.
PR      03-OCT-2000; 2000US-237606P.
PR      03-OCT-2000; 2000US-237608P.
PR      01-NOV-2000; 2000US-244867P.
PR      01-NOV-2000; 2000US-245084P.

(AVAL-) AVALON PHARM.
XX
XX      Young PE, Augustus M, Carter KC, Ebner R, Endress G, Horrigan S;
XX      Soppet DR, Weaver Z;
XX
XX      WPI; 2002-188264/24.
XX
XX      Screening for anti-neoplastic agent involves exposing cells to a
XX      chemical agent to be tested for anti-neoplastic activity, and
XX      determining a change in expression of a gene of a signature gene set
XX
XX      Claim 1; SEQ ID 6838; 44pp; English.
XX
XX      The present invention describes a method (M1) for screening for an
XX      anti-neoplastic agent. The method involves exposing cells to a chemical
XX      agent to be tested for anti-neoplastic activity, determining a change in
XX      expression of at least one gene (I) of a signature gene set, where (I)
XX      comprises a sequence (S) selected from 8447 sequences (given in ABL6164
XX      to ABL70110), or is at least 95% identical to (S), where a change in
XX      expression is indicative of anti-neoplastic activity. (I) has cytotoxic
XX      activity and can be used in gene therapy. M1 can be used for screening
XX      an anti-neoplastic agent, and can be used for producing a product which
XX      is the data collected with respect to the anti-neoplastic agent as a
XX      result of M1, and the data is sufficient to convey the chemical
XX      structure and/or properties of the agent. M1 can be used in the
XX      treatment of cancer such as colon, breast, stomach, lung, thyroid,
XX      oesophageal, ovarian, kidney, prostate or pancreatic cancer.
XX      adenocarcinoma, carcinoma, clear cell cancer, infiltrating ductal cancer,
XX      infiltrating lobular cancer, squamous cell carcinoma, neuroendocrine
XX      carcinoma, papillary carcinoma and Wilms tumour.
XX
XX      Sequence 1855 BP; 571 A; 337 C; 394 G; 553 T; 0 other;
XX
XX      Query Match      75.9%; Score 1252.6; DB 24; Length 1855;
XX      Best Local Similarity 86.1%; Pred. No. 0;
XX      Matches 1387; Conservative 0; Mismatches 224; Indels 0; Gaps 0;

```

```

Qy 21 TGCATGATCAGAGATGCTATGAATGAGCTTCACTCTCTGATACAGCTGAGCT 80
Db 1 TGCATGACACAGAGATGCTGTAATGAGCTTCACTATTTTGTATTAACATGAGCT 60
Qy 81 GTTACTTATGCTTGGGAGTTTGGAAAGTGTGTGTGGCCACAGAAATTCAGCACT 140
Db 61 TTTCCTTATGCTTGGGAAATGTGGAAAGTGTGTGTGGCCAGAAATTCAGCACT 120
Qy 141 GGATGAATATTAACATCTCGATGAACTTTGTCCAGAGAGTTCATGAGTGTAT 200
Db 121 GGATGAATATTAACATCTCGATGAACTTTGTCCAGAGAGTTCATGAGTGTAT 180
Qy 201 TGGCATCTTCACTTCCATTTCTTTGATCCCAACAGCCCATCTACTTAAATTTTGA 260
Db 181 TGGCATCTTCACTTCCATTTCTTTGATCCCAACATCTCCGCTTTAAATTTGAA 240
Qy 261 TTATCTCTATCTTTAACTAAACCTGATTTGAGATATTTCAAGCAGCTGTAGA 320
Db 241 TTATCTCCACATCTTTAACTAAACCTGATTTGAGATATTTCAATGCAAGATTA 300
Qy 321 GATGGCAGAACTTCCAAAGACACATTTTGTGATATTTTTCAGAGTACAAGAAAT 380
Db 301 GATGGCAGAACTTCCAAAGACACATTTTGTGATATTTTTCAGAGTACAAGAAAT 360
Qy 381 TGTGACATTTATGACATCTTAAAGATTTGTAAGGATATGTTCAATTAAGAAAC 440
Db 361 TGTGACATTTATGACATCTTAAAGATTTGTAAGGATATGTTCAATTAAGAAAT 420
Qy 441 TTATGAAGAACTACAGAGATCAAGATTTGATGTTTCTTGAAGTGTGTTCCT 500
Db 421 TTATGAAGAACTACAGAGATCAAGATTTGATGTTTCTTGAAGTGTGTTCCT 480
Qy 501 TTGTGAGCTGTGCGGAGTTACTTAAATACCTTTGTCTACAGCTCCGCTTCTCT 560
Db 481 GTAGTGAAGCTGTGCGGAGTTACTTAAATACCTTTGTCTACAGCTCCGCTTCTCT 540
Qy 561 CTGGCTACGCAATGAAAGCATAGTGGAGACTTCTGCTCCCTTCCATATGAGCTG 620
Db 541 CTGGCTACGCAATGAAAGCATAGTGGAGACTTCTGCTCCCTTCCATATGAGCTG 600
Qy 621 TTGTATGTGAGAACTAAGTACCAATGACTTTCAATGAGAGGATTAATAATGAT 680
Db 601 TTGTATGTGAGAACTAAGTACCAATGACTTTCAATGAGAGGATTAATAATGAT 660
Qy 681 ATGTGCTTATTTGAAATTTTGTCCAAATATTTGACATGAAGAGTGGATCACTT 740
Db 661 ATGTGCTTATTTGAAATTTTGTGTCCAAATATTTGACATGAAGAGTGGATCACTT 720
Qy 741 ACAGTGAAGTCTGAGAAAGCACTACGTTATCTGAGACAAATGSCAAAGCTGACAT 800
Db 721 ATAGTGAAGTCTGAGAAAGCACTACGTTATCTGAGACAAATGSCAAAGCTGACAT 780
Qy 801 GGCTTATTCGAATCTAGGATTTTCAATTTCTCTCAACCTTACCAATTTTGAAT 860
Db 781 GGCTTATTCGAATCTAGGATTTTCAATTTCTCTCAACCTTACCAATTTTGAAT 840
Qy 861 TCGTTGAGAGCTCACTGCAACCTGCAAAACCTTACCGAGAGAAATGGAAGATTG 920
Db 841 TTTGTTGAGAGCTCACTGCAACCTGCAAAACCTTACCGAGAGAAATGGAAGATTG 900
Qy 921 TCCAGAGCTCTGAGAAATGAGTGTGTGTGTTTCTCTGAGGTCATGTCGTAAACA 980
Db 901 TACAGAGCTCTGAGAAATGAGTGTGTGTGTTTCTCTGAGGTCATGTCGTAAACA 960
Qy 981 CGTCAGAGAAAGGCGCAATGTAATGATCAGCCCTTGCAGAGATCCCAAAAGGTT 1040
Db 961 TGACAGAGAAAGGCGCAATGTAATGATCAGCCCTTGCAGAGATCCCAAAAGGTT 1020
Qy 1041 TGTGAGATTTGATGGGATTAACCAAGATCTTAAAGATCAATCTGCTGTAAAGT 1100
Db 1021 TGTGAGATTTGATGGGATTAACCAAGATCTTAAAGATCTGCTGTAAAGT 1080

```

```

Qy 1101 GGATACCCAGAAATGATCTTCTGTCATCCCAAAACCAAGCTTTTATCATCATG 1160
Db 1081 GGATACCCAGAAATGATCTTCTGTCATCCCAAAACCAAGCTTTTATTAATCATG 1140
Qy 1161 GAATGAATGGGATCTATGAAGCTATTTTACATGGGGTCCCTATGTGGAGTTCCAT 1220
Db 1141 GAGCAATGGGATCTATGAAGCTATTTTACATGGGGTCCCTATGTGGAGTTCCAT 1200
Qy 1221 TTGGTATTCAGCTTGAATTAATAGTACATCAATGAAGGCAAGAGAGCTGTAGA 1280
Db 1201 TTGGTATTCAGCTTGAATTAATAGTACATCAATGAAGGCAAGAGAGCTGTAGA 1260
Qy 1281 ACTTCAAAATATGACAAAGGAGATTTTCTGAGGCTTTGAGAAACAGTATTCAG 1340
Db 1261 ACTTCAAAATATGACAAAGGAGATTTTCTGAGGCTTTGAGAAACAGTATTCAG 1320
Qy 1341 CTTCTTAAAGAGATGCTATGAGATTTTCAAGAAATTCACATGATCAACCTGTAA 1400
Db 1321 CTTCTTAAAGAGATGCTATGAGATTTTCAAGAAATTCACATGATCAACCTGTAA 1380
Qy 1401 CCTAGATGAGAGCTTCTGATGAGTGTGTGATGAGGCGCCCAAGAGCCAGAGAC 1460
Db 1381 CCTAGATGAGAGCTTCTGATGAGTGTGTGATGAGGCGCCCAAGAGCCAGAGAC 1440
Qy 1461 TGGGATCAGCTGCGCATGACCTCACCTGTTCCAGACACTCTATGATGTGATGG 1520
Db 1441 TTGGGATCAGCTGCGCATGACCTCACCTGTTCCAGACACTCTTGTGATGTGATGG 1500
Qy 1521 TCTGTGACCTGTGTGGCAATGCTATATCTTGTTCACAAATGTTTATTTTCT 1580
Db 1501 TCTGTGACCTGTGTGGCAATGCTATATCTTGTTCACAAATGTTTATTTTCT 1560
Qy 1581 GTCAAAATTTATTAATACTGAAGATGAGAAAGAGGAAATGATCTTTC 1631
Db 1561 TCTGAAGTGTGCTAGAAAGCAAGAAAGGAAATGATGATTAATC 1611

RESULT 9
ABL68868
ID ABL68868 standard, DNA, 1855 BP.
XX
AC ABL68868;
XX
DT 15-MAY-2002 (first entry)
XX
DE Kidney cancer related gene sequence SEQ ID NO:7205.
XX
KW Human; cancer; colon; breast; ovary; oesophagus; kidney; thyroid;
KW stomach; lung; prostate; pancreas; carcinoma; antitumour; cancerous;
KW cytostatic; gene therapy; antineoplastic; Wilms' tumour; adenocarcinoma;
KW gene; ds.
XX
OS Homo sapiens.
XX
PN W0200194629-A2.
XX
PD 13-DEC-2001.
XX
PF 30-MAY-2001; 2001MO-US10838.
XX
PR 05-JUN-2000; 2000US-209473P.
PR 05-JUN-2000; 2000US-209531P.
PR 18-SEP-2000; 2000US-233133P.
PR 18-SEP-2000; 2000US-233617P.
PR 20-SEP-2000; 2000US-234009P.
PR 20-SEP-2000; 2000US-234034P.
PR 20-SEP-2000; 2000US-234052P.
PR 22-SEP-2000; 2000US-234509P.
PR 22-SEP-2000; 2000US-234567P.
PR 25-SEP-2000; 2000US-234923P.
PR 25-SEP-2000; 2000US-234924P.
PR 25-SEP-2000; 2000US-235077P.
PR 25-SEP-2000; 2000US-235082P.

```

PR 25-SEP-2000; 2000US-235134P.
 PR 25-SEP-2000; 2000US-235280P.
 PR 26-SEP-2000; 2000US-23537P.
 PR 26-SEP-2000; 2000US-235638P.
 PR 27-SEP-2000; 2000US-235711P.
 PR 27-SEP-2000; 2000US-235720P.
 PR 27-SEP-2000; 2000US-235840P.
 PR 27-SEP-2000; 2000US-235863P.
 PR 28-SEP-2000; 2000US-236028P.
 PR 28-SEP-2000; 2000US-236032P.
 PR 28-SEP-2000; 2000US-236033P.
 PR 28-SEP-2000; 2000US-236034P.
 PR 28-SEP-2000; 2000US-236109P.
 PR 28-SEP-2000; 2000US-236111P.
 PR 29-SEP-2000; 2000US-236842P.
 PR 29-SEP-2000; 2000US-236891P.
 PR 02-OCT-2000; 2000US-237172P.
 PR 02-OCT-2000; 2000US-237173P.
 PR 02-OCT-2000; 2000US-237278P.
 PR 02-OCT-2000; 2000US-237294P.
 PR 02-OCT-2000; 2000US-237295P.
 PR 02-OCT-2000; 2000US-237316P.
 PR 03-OCT-2000; 2000US-237425P.
 PR 03-OCT-2000; 2000US-237598P.
 PR 03-OCT-2000; 2000US-237604P.
 PR 03-OCT-2000; 2000US-237606P.
 PR 03-OCT-2000; 2000US-237608P.
 PR 01-NOV-2000; 2000US-244867P.
 PR 01-NOV-2000; 2000US-245084P.
 XX
 PA (AVALON PHARM.
 XX
 PI Young PE, Augustus M, Carter KC, Ebner R, Endress G, Horrigan S;
 PI Soppet DR, Weaver Z;
 XX
 XX WPI, 2002-188264/24.
 XX
 PT Screening for anti-neoplastic agent involves exposing cells to a
 PT chemical agent to be tested for anti-neoplastic activity, and
 PT determining a change in expression of a gene of a signature gene set -
 XX
 XX
 PS Claim 1; SEQ ID 7205; 44p; English.

CC The present invention describes a method (M1) for screening for an
 CC anti-neoplastic agent. The method involves exposing cells to a chemical
 CC agent to be tested for anti-neoplastic activity, determining a change in
 CC expression of at least one gene (I) of a signature gene set, where (I)
 CC comprises a sequence (S) selected from 8447 sequences (given in ABL61664
 CC to ABL70110), or is at least 95% identical to (S), where a change in
 CC expression is indicative of anti-neoplastic activity. (I) has cytostatic
 CC activity and can be used in gene therapy. M1 can be used for screening
 CC an anti-neoplastic agent, and can be used for producing a product which
 CC is the data collected with respect to the anti-neoplastic agent as a
 CC result of M1, and the data is sufficient to convey the chemical
 CC structure and/or properties of the agent. M1 can be used in the
 CC treatment of cancer such as colon, breast, stomach, lung, thyroid,
 CC oesophageal, ovarian, kidney, prostate or pancreatic cancer,
 CC adenocarcinoma, carcinoma, clear cell cancer, infiltrating ductal cancer,
 CC infiltrating lobular cancer, squamous cell carcinoma, neuroendocrine
 CC carcinoma, papillary carcinoma and Wilms' tumour.

CC
 XX Sequence 1855 BP; 571 A; 337 C; 394 G; 553 T; 0 other;

Query Match 75.9%; Score 1252.6; DB 24; Length 1855;
 Best Local Similarity 86.1%; Pred. No. 0;
 Matches 187; Conservative 0; Mismatches 224; Indels 0; Gaps 0;

QY 21 TGCATTGCTCAGATGTCTATGAATGACCTTCAGCTTCTCCGATAGAGCTGAGT 80
 DB 1 TGCATTGCTCAGATGTCTATGAATGACCTTCAGCTTCTCCGATAGAGCTGAGT 60
 QY 81 GTTACTTGTAGCTGTGGAGTGTGGAAAGGTGCTGGTGGCCCAAGAACTTCAGCACT 140
 DB 1 GTTACTTGTAGCTGTGGAGTGTGGAAAGGTGCTGGTGGCCCAAGAACTTCAGCACT 140

DB 61 TTGCTTTAGCTGTGGGAATTGTGAAAGGTGCTGTGTGGGAGCAAGATACAGCCATT 120
 QY 141 GGAATGATATTAAGACAATCCCTGATGAACTGTGCCAGAGAGGTCATGAGTACTAT 200
 DB 121 GGAATGATATTAAGACAATCCCTGATGAACTGTGCCAGAGAGGTCATGAGTACTAT 180
 QY 201 TGGCATCTTCAGCTTCATTTCTTGTGATCCCAAGCCCATCTACTTTAAATTTGAAG 260
 DB 181 TGGCATCTTCAGCTTCATTTCTTGTGATCCCAAGCCCATCTACTTTAAATTTGAAG 240
 QY 261 TTTATCTGTATCTTTTACTTAAACGAGTTTGAGATATTTATCAGACCTGTTAAGA 320
 DB 241 TTTATCCCATCATCTTTTAACTTAAACGAGTTTGAGATATTTATCAGACCTGTTAAGA 300
 QY 321 GATGGGAGACCTTCCAAAGACATTTTGTGATATTTTTCACAAAGTACAAAGATCA 380
 DB 301 GATGGTACACCTTCCAAAGACATTTTGTGATATTTTTCACAAAGTACAAAGATCA 360
 QY 381 TGTGACATTTTAAATGACATCTTAAAGATTTGTAGATATTTAGTTTCAATTAAGAAC 440
 DB 361 TGTCAATTTTGTGACATCTTAAAGATTTGTAGATATTTAGTTTCAATTAAGAAC 420
 QY 441 TTTATGAAAGAACTACAGAGAGTCAAGATTTGATGTTCTTTCAGAGAGCTGTTCCCT 500
 DB 421 TTTATGAAAGAAAGTACAGAGAGTCAAGATTTGACGCTTTTTCAGAGAGCTGTTCCCT 480
 QY 501 TTGGTGAAGCTGTGGGAGGCTTCTTAAATACCTTGTGTCTACAGCTCGCTTCTCTC 560
 DB 481 GTATGAGCTGTGGGAGGCTTCTTAAATACCTTGTGTCTACAGCTCGCTTCTCTC 540
 QY 561 CTGGCTACGCAATTTGAAAGCATATGAGAGAGCTTCTGTTCCCTCTTCTTANGGCTCG 620
 DB 541 CTGGCTACGCTTGTGAAAGCATATGAGAGAGCTTCTGTTCCCTCTTCTTANGGCTCG 600
 QY 621 TTTGATGTCAGAACTTAAAGTACCAATGACTTTTCAATGAGAGGTTAAATATGATCT 680
 DB 601 TTTGATGTCAGAACTTAAAGTACCAATGACTTTTCAATGAGAGGTTAAATATGATCT 660
 QY 681 ATGAGCTTTTATTTGAAATTTTGTGCAATATTTGACATGAAAGGAGTCACTTCT 740
 DB 661 ATGAGCTTTTATTTGAAATTTTGTGCAATATTTGACATGAAAGGAGTCACTTCT 720
 QY 741 ACAGTGAAGTCTTGAAGAGACCCACTACGTTATCTGAGACAAATGGCAAAAGCTGACATAT 800
 DB 721 ATAGTGAAGTCTTGAAGAGACCCACTACGTTATCTGAGACAAATGGCAAAAGCTGACATAT 780
 QY 801 GCGTTATTTGCAACTTACCTGGATTTTCAATTTCTCTACCCACTTTTACCAATTTGAT 860
 DB 781 GCGTTATTTGCAACTTACCTGGATTTTCAATTTCTCTACCCACTTTTACCAATTTGAT 840
 QY 861 TCGTTGAGAGCTCCACTGCAAACTGCAAAACCCCTACCGAAGGAATGGAAGATTG 920
 DB 841 TTTGTTGAGAGCTCCACTGCAAACTGCAAAACCCCTGCTTACGAAATGGAAGATTG 900
 QY 921 TCCAGAGCTCTGAGAAATGTTGTGTGTTTCTCTGAGGTCATGTGTCAGTACA 980
 DB 901 TACAGAGCTCTGAGAAATGTTGTGTGTTTCTCTGAGGTCATGTGTCAGTACA 960
 QY 981 CGTACAGAAAGAGCCCAATGTTATTTGATATAGCCCTTCCCAAGATCCCAAAAGGTTTC 1040
 DB 961 TGACAGAAAGAGCCCAATGTTATTTGATATAGCCCTTCCCAAGATCCCAAAAGGTTTC 1020
 QY 1041 TGTGAGATTTGATGAGGAATTAACAGATCTTTAGAGCTCAATATCTGCTGTACAAGT 1100
 DB 1021 TGTGAGATTTGATGAGGAATTAACAGATCTTTAGAGCTCAATATCTGCTGTACAAGT 1080
 QY 1101 GGATACCCCAAGATCTTCTGTGATCCCAAAACCAAGCTTTTATCACTGATGTTG 1160
 DB 1081 GGATACCCCAAGATCTTCTGTGATCCCAAAACCAAGCTTTTATCACTGATGTTG 1140
 QY 1161 GAATGATGAGATCTATGAGCTATTTTCAATGAGGTTCCCTATGTTGGAGATTTCCCATAT 1220
 DB 1141 GAGCAATGAGATCTATGAGCTATTTTCAATGAGGTTCCCTATGTTGGAGATTTCCCATAT 1200


```

Qy 506 GAGCTGCTGCCGAGTACTTAAATAACCTTTGTCTACAGCTCCGCTCTCTCTG 565
Db 481 GAGCTGCTGCCGAGTACTTAAATAACCTTTGTCTACAGCTCCGCTCTCTCTG 540
Qy 566 TACCGAATTGAAAGCAATAGTGGAGACTTCTGTTCCCTCCTTATGTCCTGTTT 625
Db 541 TACCAATTTGAAAGCAGAGTGGAGACTGATTTTCCCTCCTTACATCTATGTT 600
Qy 626 ATGTCAGAACTAAGTGAACCAATGACTTTCATAGAGAGGTAAATAATGATCTATG 685
Db 601 ATGTCAGAACTAAGTGAACCAATGACTTTCATAGAGAGGTAAATAATGATCTATG 660
Qy 686 CTTTATTTTGAATTTGGTCCAAATATTTTGCATGAGAAAGTGGATCACTTCACT 745
Db 661 CTTTATTTTGAATTTGGTCCAAATGCTGTGATAGAAAGAGTGGATCACTTCACT 720
Qy 746 GAAGTTCTAGAGAAAGCCCACTACGTTATCTGAGCAATGACAAAAGTGAATAGGCTT 805
Db 721 GAAGTTTAGAGAAAGCCCACTACCTTATTTGAGCAATGAGAAAGTGAATAGGCTT 780
Qy 806 ATTGGAATCTAGTGGATTTTTCATTTCTCAACCACCTTACCAATGTTGAGTTGTT 865
Db 781 ATGCAAACTCTGAGTTTTCATTTCTCATTCATTCATCAAAAGTTGATTTTGT 840
Qy 866 GGAGGACTCCACTGCAACTGCAAAACCCCTACCGAAGAAATGAGAGTTGTCAG 925
Db 841 GGAGGATTCCTCACTGCAAACTGCAAAACCCCTACCTAAGAAATGAGAGTTGTCAG 900
Qy 926 AGCTCTGAGAAATAGTGTGTGTGTGTGTCTCTGAGGATCGATGATGATGATGATG 985
Db 901 AGCTCTGAGAAATAGTGTGTGTGTGTGTCTCTGAGGATCGATGATGATGATGATG 960
Qy 986 GAAGAAAGGCGCAATGTAATGTCATCAGCCCTTGGCAAGATCCCAAAAGTTCTGTG 1045
Db 961 GCAAGAAAGGCGCAATGTAATGTCACAGCCCTTGGCAAGATCCCAAAAGTTCTGTG 1020
Qy 1046 AGATTGATGGAAATPAAACAGATCTTATAGACTCAATACCTGGCTGTACAGTGTGATA 1105
Db 1021 AGATTGATGGAAATPAAACAGATCTTATAGACTCAATACCTGGCTGTACAGTGTGATA 1080
Qy 1106 CCCGAGATGATCTTCTGTGATCCCAAAACCAAGCTTTTATCACTCATGTGAGATG 1165
Db 1081 CCCGAGATGATCTTCTGTGATCCCAAAACCAAGCTTTTATCACTCATGTGAGATG 1140
Qy 1166 AATGGATCTATGAAAGCTATTTACATGGGCTCTTATGTTGGAGTTCCCATATTTGTT 1225
Db 1141 AATGGATCTATGAAAGCTATTTACATGGGCTCTTATGTTGGAGTTCCCATATTTGTT 1200
Qy 1226 GATCAGCTGATTAATCATAGCTCAGATGAAAGCCAAAGAGAGCTGTGAATAAATTC 1285
Db 1201 GATCAGCTGATTAATCATAGCTCAGATGAAAGCCAAAGAGAGCTGTGAATAAATTC 1260
Qy 1286 AAAAATATGACAAAGGAAATTTACTGAGGCTTTGAGAACAGTCAATCCGATTCCTCT 1345
Db 1261 AAAAATATGACAAAGGAAATTTACTGAGGCTTTGAGAACAGTCAATCCGATTCCTCT 1320
Qy 1346 TATTAAGAAATGCTATGATATTAACAAGATTCACATGATCAACTGTAAAGCCCTTA 1405
Db 1321 TATTAAGAAATGCTATGATATTAACAAGATTCACATGATCAACTGTAAAGCCCTTA 1380
Qy 1406 GATGAGAGATCTTCTGATGAGTTGTCTCATGCGCCCAAGAGAGCCCAACACTGTGA 1465
Db 1381 GATGAGAGATCTTCTGATGAGTTGTCTCATGCGCCCAAGAGAGCCCAACACTGTGA 1440
Qy 1466 TCAGCTGCCAATGACCTCACTGTTTCCAGACTCTATAGATGTTGAGTTGCTG 1525
Db 1441 GTTGAGGCGCAATGACCTCACTGTTTCCAGACTCTATGATGTTGAGTTGCTG 1500
Qy 1526 CTGACCTGTGTGGCACTGCTATATTTCTTGTTCACAAATGTTTATTTTCTGTCAA 1585
Db 1501 CTGACCTGTGTGGCACTGCTATATTTCTTGTTCACAAATGTTTATTTTCTGTCAA 1560

```

```

Qy 1586 AATTTAATAAACTAGAAAGATGAAAGAGGAGGATGATCTTTC 1631
Db 1561 AAGTTTCTGTAAGAAAGGAAAGAGGAGGAAAGAAATTAATGTTATCTC 1606

RESULT 12
ABN97261
ID ABN97261 standard; DNA; 2799 BP.
XX
XX AC ABN97261;
XX
XX DT 13-AUG-2002 (first entry)
XX
XX DE Gene #3759 used to diagnose liver cancer.
XX
XX KM Gene; liver cancer; ds; hepatocellular carcinoma; hepatocytic;
XX KM metastatic liver tumour; cytotoxic; expression profile; disease state;
XX KM disease progression; drug toxicity; drug efficacy; drug metabolism.
XX OS Homo sapiens.
XX
XX EN W0200229103-A2.
XX
XX PD 11-Apr-2002.
XX
XX PF 02-OCT-2001; 2001WO-US30589.
XX
XX PR 02-OCT-2000; 2000US-237054P.
XX
XX PA (GENE-) GENE LOGIC INC.
XX
XX PI Horne D, Alvares C, Peres-Da-Silva S, Vockley JG;
XX
XX WP1; 2002-426119/45.
XX
XX DR Diagnosing and detecting the progression of liver cancer,
XX PT hepatocellular carcinoma or metastatic liver tumor in a patient,
XX PT involves detecting the level of expression of two or more genes in a
XX PT liver tissue sample -
XX
XX CC Claim 1; SEQ ID NO 3759; 298bp; English.
XX
XX PS The invention relates to a novel method for diagnosing and detecting the
XX CC progression of liver cancer, hepatocellular carcinoma or metastatic liver
XX CC tumour in a patient, and differentiating metastatic liver cancer from
XX CC hepatocellular carcinoma in a patient, involving detecting the level of
XX CC expression of two or more genes represented in ABN93503-ABN97455 in a
XX CC tissue sample. The method of the invention has hepatocytic, and
XX CC cytosolic activity. The method is useful for diagnosing and detecting
XX CC the progression of liver cancer, hepatocellular carcinoma and metastatic
XX CC liver carcinoma in a patient. The method is useful for identifying
XX CC expression profiles which serve as useful diagnostic markers as well as
XX CC markers that can be used to monitor disease states, disease progression,
XX CC drug toxicity, drug efficacy and drug metabolism.
XX CC Note: The sequence data for this patent did not form part of the printed
XX CC specification, but was obtained in electronic format directly from WIPO
XX CC at ftp.wipo.int/pub/published_pat_sequences.
XX
XX SQ Sequence 2799 BP; 792 A; 605 C; 569 G; 833 T; 0 other;

Query Match 74.3%; Score 1225.2; DB 24; Length 2799;
Best Local Similarity 85.6%; Pred. No. 0;
Matches 1375; Conservative 0; Mismatches 228; Indels 3; Gaps 1;

Qy 26 TGCATCAGATGCTATGAATGAGCTTCACTTCTCTCTGATACAGCTGAGCTGTTAC 85
Db 2 TGCAGAAAGATGGCTGGAATGAGACTACAG---TTCTGCTGATACAACTCAAGTTTAC 58
Qy 86 TTTAGCTCTGGAGTTGTGAAAAGTGTCTGTGTGGCCCAAGAAATTCAGCCTGATG 145
Db 59 TTTAGCTCTGGAGTTGTGAAAAGTGTGTGTGTGGCCCAAGAAATTCAGCCTTGTGATG 118
Qy 146 AATTAAGAAATCTCGATGAACTGTGTCAGAGAGGTATGAGTGAATGATGCA 205

```



```

Db 119 AATATGAAGACATCTGAAAGAACTGTTGAGAGAGGTGATGAGTACGTGACGCA 178
Qy 206 TCTTCAAGTCTGATTTCTTTGATGATCCCAAGACCCTACTACTCTTAAATTTGAATTTAT 265
Db 179 TCTTCAAGTCTGATTTCTTTGATGATCCCAAGACCCTACTACTCTTAAATTTGAATTTAT 238
Qy 266 CCTGATCTTTAACTGAAAGTGTGAGATATTTCAAGACAGCTGTTAAAGATG 325
Db 239 CCTACATCTTTAACTGAAATTTGATGATATCATCATGCAATTTGTTAAAGATG 298
Qy 326 GCAGAACTTCCAAAAGACATTTTGTCTATATTTTCAACAATGCAAGAAATCATGTG 385
Db 299 TCAGAAATTTCAAAAAGATATTTTGTGTTACCTTTTCAACAAGAACAAATCCTGTG 358
Qy 386 ACATTTATGACATCTTGAAGAGTCTGTGAAGATATAGTTTCAAAATGAATTTATG 445
Db 359 GCAATTTATGACATATATGAAACCTCTGTAAAGATGATTTCAAAATGAATTTATG 418
Qy 446 AAGAACTACAGAGTCAAGATTTGATGTTGTTCTGAGATGCTGTTTCCCTTTGAT 505
Db 419 AAAAACTACAGAGTCAAGATTTGATGTTGTTCTGAGATGCTGTTTCCCTTTGAT 478
Qy 506 GAGTGTGCTGCGAGATCTTAAATACCTTTGTTCTACAGCTGCTGCTCTCTGCTG 565
Db 479 GAGTGTGCTGCGAGATCTTAAATACCTTTGTTCTGTTGTTGTTGTTGTTGTTG 538
Qy 566 TACGCAATTTGAAAGACATGAGAGAGCTCTGTTTCCCTCTCTATGCTGTTGTT 625
Db 539 TACTCATTTGAAAGGACAGTGAAGATTTATTTTCCCTCTCTCTGAGTCTGTTGTT 598
Qy 626 ATGTGAGAACTAAGTGAACCAATGACTTTCATAGAGAGGTTAAATATGATCTATG 685
Db 599 ATGTGAAATTTAAGTGAATGATGATCTTTCATGAGAGAGGTTAAATATGCTATG 658
Qy 686 CTTTATTTTGAATTTTGTGTTCCAAATATTTGACATGAAGAACTGAGTCTTACAGT 745
Db 659 CTTTATTTTGAATTTTGTGTTCCAAATATTTAATGAAAGTGGATCAGTTTACAGT 718
Qy 746 GAACTTGAAGAAAGCCCACTGCTTATCTGAGCAATGAGCAAAAGCTGATGCTT 805
Db 719 GAACTTGAAGAAAGCCCACTGCTTATCTGAGCAATGAGCAAAAGCTGATGCTT 778
Qy 806 ATTGAACTACTGAGATTTTCAATTTCTCTACCCACTCTTACCAATGTTGATGCTT 865
Db 779 ATGCAAACTCTGGAATTTTAAATTTCTCTCTCATCTTCTTACCAATGTTGATGCTT 838
Qy 866 GAGAGACTCCACTGCAAACTGCAAAACCCCTACCGAAGAAATGAAAGTTGTCAG 925
Db 839 GAGAGACTCCACTGCAAACTGCAAAACCCCTACCGAAGAAATGAAAGTTGTCAG 898
Qy 926 AGCTTCTGAGAAAGATGCTGTTGTTTCTCTGAGGCTGATGCTGATGACGTC 985
Db 899 AGCTTCTGAGAAAGATGCTGTTGTTTCTCTGAGGCTGATGCTGATGACGTC 958
Qy 986 GAAAGAAAGGCGCAATGTAATGTCATCAGCCCTGCGCAAGATCCCAAAAGGTTCTG 1045
Db 959 GAAAGAAAGGCGCAATGTAATGTCATCAGCCCTGCGCAAGATCCCAAAAGGTTCTG 1018
Qy 1046 AGATTTGATGGAATTAACAGATCTTATGAGACTCAATATCTGAGTGTACAGTGATA 1105
Db 1019 AGATTTGATGGAATTAACAGATCTTATGAGTGTCAATATCTGAGTGTACAGTGATA 1078
Qy 1106 CCCCAAGATGATCTTCTGTTGATGCTGCAAAAGCTTTTATCATCTCATGTTGAATG 1165
Db 1079 CCCCAAGATGATCTTCTGTTGATGCTGCAAAAGCTTTTATCATCTCATGTTGAATG 1138
Qy 1166 AATGAGATTAAGAGCTATTTACATGAGGCTCCATATGTTGAGGATTCCTCATTTTGT 1225
Db 1139 AATGAGATTAAGAGCTATTTACATGAGGATTCCTATGTTGAGGATTCCTCATTTTGT 1198
Qy 1226 GATCAGCTTGAATACATGCTGATGAGGCGCAAGAGGAGCTGTGAAATTAACCTTC 1285

```

```

Db 1199 GATCAACCTGATATATTTGCTCATGATGAGGCCAAGGAGACGCTGTAGAGTGACTTC 1258
Qy 1286 AAAAATATGACAGCGAAGATTTTACTGAGGCTTTGAGAACAGTATTAACGATTCCT 1345
Db 1259 AACCAATGCTGAGTACAGACCTGCTTAATGCAATGCAAGACAGTATTAATGATCTTCA 1318
Qy 1346 TATTAAGAAATGCTATGATGATTAATCAAGAAATTCACATGATCAACCTGTTAAAGCCCTTA 1405
Db 1319 TATTAAGAAATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1378
Qy 1406 GATGAGACAGCTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1465
Db 1379 GATGAGACAGCTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1438
Qy 1466 TCAGCTGCTGAGACAGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1525
Db 1439 GTTCAAGCCCAACCACTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1498
Qy 1526 CTGACCTGCTGAGACAGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1585
Db 1499 CTGCTGCTGCTGAGACAGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1558
Qy 1586 AATTTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1631
Db 1559 AAGTTGCTGAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 1604

RESULT 13
AAD06844
ID AAD06844 standard; DNA, 1713 BP.
AC
XX
AC AAD06844;
XX
DT 06-AUG-2001 (first entry)
XX
DE Human breast cancer specific gene-2 (BCSG-2) #1.
XX
KW Human; breast cancer specific gene-2; BCSG-2; cytostatic; vaccine;
XX breast cancer; therapeutic; gene therapy; ds.
XX
OS Homo sapiens.
XX
PN MO200137779-A2.
XX
PD 31-MAY-2001.
XX
PF 22-NOV-2000; 2000MO-US32056.
XX
PR 23-NOV-1999; 99US-0166973.
XX
PA (DIAD-) DIADEXUS INC.
XX
PI Salceda S, Caffery R, Recipon H, Sun Y;
XX
DR WPI, 2001-367602/38.
XX
PT Novel breast cancer specific gene for diagnosing, monitoring, staging,
PT imaging, preventing and treating cancers, particularly breast cancer -
XX
PS Claim 1; Page 49; 66pp; English.
XX
CC The invention relates human breast cancer specific genes (BCSG's) and
CC their corresponding proteins. BCSG is useful for diagnosing, staging,
CC monitoring, imaging, preventing and treating breast cancers. BCSG is also
CC useful for inducing an immune response against a target cell expressing
CC BCSG. The invention also provide methods for detecting genetic lesions or
CC mutations in BCSG, thereby determining if a human with the genetic lesion
CC is at risk for breast cancer or has breast cancer. BCSG antibodies
CC labelled with paramagnetic ions or radioisotopes is useful for imaging
CC breast cancers, while BCSG antibodies conjugated to a cytotoxic agent is
CC useful for treating breast cancer. BCSG is useful in the rational design
CC of new therapeutics for imaging and treating cancers. BCSG is also used
CC in gene therapy. The present DNA sequence is human breast cancer specific

```


CC gene-2 (BCSG-2) or Gene ID 480489.

XX Sequence 1713 BP; 507 A; 320 C; 375 G; 511 T; 0 other;

Query Match 74.0%; Score 1221.4; DB 22; Length 1713;

Best Local Similarity 85.6%; Pred. No. 0;

Matches 1382; Conservative 0; Mismatches 231; Indels 2; Gaps 2;

QY 19 ATTGCATTCATCGAGATGCTATGAATAATGAGCTTCACTCTCTCCCTGATACAGCT-GA 77
 DB 1 ATCCGATTCGACCCAGATGACTCTGAATAAGCACTTCACTCTCTGATACATCTCCA 60
 QY 78 GCTGTACTTACCTGCTGGAGTTGTGAAAAGTGTGGTGGCCCAAGATTCAGCC 137
 DB 61 GTTGTTACTTACCTGCTGGAGTTGTGAAAAGTGTGGTGGCCCAAGATTCAGCC 120
 QY 138 ACTGATGAATATTAAGCAATCTCGATGAATCTTGTCCAGAGAGGTATAGGTACTG 197
 DB 121 ATTGATGCAATATTAAGCAATCTCGAAGAGCTTGTTCAGAGAGGTATAGGTACTG 180
 QY 198 TATGGGATCTTCAGCTTCCATTTCTTTCGATCCCAAGCCCATCTACTTAAATTTG 257
 DB 181 TACTGGCACTTCAGCTTCCATTTCTTTCGATCCCAAGATGATGCACTCTTAAATTTG 240
 QY 258 AAGTTTATCTGTATCTTTAACTAAACTGAGTTGAGATATTTATCAAGCAGCTGTTA 317
 DB 241 AAGTTTATCTGTATCTTTAACTAAACTGAGTTGAGATATTTATCAAGCAGCTGTTA 300
 QY 318 AGAGATGGGCAAACTTCCAAAAGACACATTTGGTCAATATTTTTCACAGTCAAGAAA 377
 DB 301 AGAATGCTGACATCTGAAAAGATAGCTTTGGTATATTTTTCACAGTCAAGAAA 360
 QY 378 TCATGTGGACCTTTAATGACATCTTAAAGTCTGTAAGGATATGTTCAATATAGA 437
 DB 361 TCCGTGGGAATTAATGACATCTTAAAGTCTGTAAGGATATGTTCAATATAGA 420
 QY 438 AACTTATGAAGAACTACAGAGTCAAGATTTGATTTGTTCTTGACAGTCTGTTTCC 497
 DB 421 AAGTTATGAAGAACTACAGAGTCAAGATTTGATTTGATTTGTTCTTGACAGTCTGTTTTC 480
 QY 498 CTTTGGTACCTGCTGGCCGAGTTACTTAAATATCCCTTGTCTACAGCTCCGCTTCT 557
 DB 481 CTTGTTGGAGCTGCTGGCTCGCTACTTAATACGCTTGTGTACAGTCTCCGCTTTA 540
 QY 558 CTCCTGGTACGCAATGAAGAAAGCATAGTGGAGACTTCTCCCTCCCTCATATGTC 617
 DB 541 CTCCTGGTACGCAATGAAGAAAGCATAGTGGAGACTTCTCCCTCCCTCATATGTC 600
 QY 618 CTGTTGTTATGTCAGAACTAAGTACCAATGACCTTTCATGAGAGGGTAAATAATAGA 677
 DB 601 CTATTTGTTATGTCAGAACTAAGTACCAATGACCTTTCATGAGAGGGTAAATAATAGA 660
 QY 678 TCTATGCTTATTTTGAATTTTGGTTCCAAATATTTGACATGAAGAAGTGGATCAGT 737
 DB 661 TCTATGCTTATTTTGAATTTTGGTTCCAAATGTCGATATGAAGAAGTGGATCAGT 720
 QY 738 TCTACAGTGAAGTCTGAGAGACCCCACTAGTATTCAGAGACAAATGGCAAAAGCTGACA 797
 DB 721 TTTACAGTGAAGTCTGAGAGACCCCACTAGTATTTGAGACAAATGGCAAAAGCTGACA 780
 QY 798 TATGGCTTATTCGAAGAACTACTGAGATTTTCAATTTCCACCCACTTTACCAATGTTG 857
 DB 781 TATGGCTTATTCGAAGAACTACTGAGATTTTCAATTTCCATCCATCTTTACCAAGCTTG 840
 QY 858 AGTTGCTTGAAGACTCCACT-GCAAACTGCAAAAGCCCTACCGAAGAAATGAAAGAG 916
 DB 841 ATTGTTGTTGAGGATTCACATGCAAAAGCTCCAAAGCCCTACCGAAGAAATGAAAGAG 900
 QY 917 TTTGTCAGAGCTCTGAGAGAAATGAGTGTGTTTCTCTGGGGTCAATGCTCAGT 976
 DB 901 TTTGTCAGAGCTCTGAGAGAAATGAGTGTGTTTCTCTGGGGTCAATGCTCAGT 960
 QY 977 AACACGTCAGAAAGGAGCAATGATATGATCAGCCCTTGCAGAGATCCCAAAAG 1036

DB 961 AACATGACAGAGAAAGGCCCAATGTAATTGCAACAGCCCTTGGCAAGATCCCAAAAG 1020
 QY 1037 GTTCTGAGAGATTTGATGGGAATTAACCAAGATATCTTTAGAGCTCAATACTGGCTGAC 1096
 DB 1021 GTTCTGAGAGATTTGATGGGAATTAACCAAGATATCTTTAGAGCTCAATACTGGCTGAT 1080
 QY 1097 AAGTGAATACCCAGAAATGATCTTCTGTGATCCCAAAACCAAGGCTTTATCACTCAT 1156
 DB 1081 AAGTGAATACCCAGAAATGATCTTCTGTGATCCCAAAACCAAGGCTTTATCACTCAT 1140
 QY 1157 GGTGGAATGAATGGGATCTATGAAGCTATTTACCATGGGGTCCCTATGTTGGAGTTCC 1216
 DB 1141 GGTGGAATGAATGGGATCTATGAAGCTATTTACCATGGGGTCCCTATGTTGGAGTTCCA 1200
 QY 1217 ATATTTGATATCAGCTTATTAATCATAGTCTCAATGAAGGCCAAAGAGACAGCTGTAGA 1276
 DB 1201 TGTGTTGGGATCAACCTGATTAACATTTGCTCAATGAAGGCCAAAGAGAGAGCTGTAGA 1260
 QY 1277 ATAACTTCAAACTATGACAAAGCAAGATTTACTGAGGCTTTGAGAAAGTCAATACC 1336
 DB 1261 TTGGACTTCAACCAATGTGAGTACAGACCTGCTGAATGCTACTGAAGACATTAAT 1320
 QY 1337 GATTCCTTTATTAAGAAATGCTATGATTAATCAAGAAATTCACATGATCAACCTGTA 1396
 DB 1321 GATTCCTTTATTAAGAAATGCTATGATTAATCAAGAAATTCACATGATCAACAGTA 1380
 QY 1397 AAGCCCTATATGAGAGAGCTTCTGATGATGATTTGTCATGGGCCCAAGAGCCAG 1456
 DB 1381 AAGCCCTATATGAGAGAGCTTCTGATGATGATTTGTCATGGGCCCAAGAGCCCAA 1440
 QY 1457 CACCTGGATCAGCTGGCCCATGACCTGACCTGGTCCAGCACTACTATAGATGAT 1516
 DB 1441 CACCTGGATCAGCTGGCCCATGACCTGACCTGGTCCAGCACTACTATAGATGAT 1500
 QY 1517 GGGTTCCTGCTGACCTGTGTGGCAACTGTATATTTCTTGTTCACAAATGTTTTTATTT 1576
 DB 1501 GGGTTCCTGCTGACCTGTGTGGCAACTGTATATTTATCATCAAAAGTTTGTCTGTTT 1560
 QY 1577 TCTGTCAAAATTTTATTAATACTAGAAAGATGAAAAGGGAGATGATCTTTC 1631
 DB 1561 TGTTCGTGAAGTTTGTGTAGAAAAGGAGAAAGGAGAAAGAGATTATGATGTC 1615

RESULT 14
 ID AAS69710 standard; cDNA; 1859 BP.
 XX AAS69710;
 AC
 XX
 DT 13-FEB-2002 (first entry)
 XX
 DE DNA encoding novel human diagnostic protein #5514.
 XX
 KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
 XX food supplement; medical imaging; diagnostic; genetic disorder; ss.
 OS Homo sapiens.
 XX
 PN W0200175067-A2.
 PD 11-OCT-2001.
 XX
 PF 30-MAR-2001; 2001MO-US08631.
 XX
 PR 31-MAR-2000; 2000US-0540217.
 XX 23-AUG-2000; 2000US-0649167.
 PA (HXS-) HXSQ INC.
 XX
 PI Dmanac RT, Liu C, Tang YT;
 XX
 DR WPI; 2001-639362/73.

DR P-PSDB; ABG05523.
XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity
XX
PS Claim 1; SEQ ID No 5514; 103bp; English.
XX
CC The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantifying a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAS64197-AAS94564 represent novel human
CC diagnostic coding sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 1859 BP; 570 A; 337 C; 397 G; 555 T; 0 other:
Query Match 73.2%; Score 1207; DB 23; Length 1859;
Best Local Similarity 85.8%; Pred. No. 0;
Matches 1386; Conservative 0; Mismatches 225; Indels 4; Gaps 4;
QY 21 TGCATTCGATCAGAGTGTCTATGAATGAGACTTCTCTCGATACAGTACGT 80
DB 1 TGCATTCGACCGAGATGTCTGAAATGACTTCAATTTTGTAAATACACTGAGCT 60
QY 81 GTTACTTTAGCTCTGGAGTTGTGAAAGTGCTGGTGGCCACAGAAATTCAGCCACT 140
DB 61 TTTGCTTTAGCTCTGGAGTTGTGAAAGTGCTGGTGGCCACAGAAATTCAGCCACT 120
QY 141 GGATGAATATTAAGAACAATCTCGATGAGCTTATTCAGAGAGTCATGAGTCACTGAT 200
DB 121 GGATGAATATTAAGAACAATCTCGATGAGCTTATTCAGAGAGTCATGAGTCACTGAT 180
QY 201 TGGCATCTTCAGCTCCATCTTCTTGATCCCAAGCCCATCTCTTAAATTTGAAG 260
DB 181 TGGCATCTTCAGCTCCATCTTCTTGATCCCAAGCCCATCTCTTAAATTTGAAG 240
QY 261 TTTATCTGTATCTTTAACTAAAGTGTGAGATTTATTCAGAGAGTGTGTTAAGA 320
DB 241 TTTATCTGTATCTTTAACTAAAGTGTGAGATTTATTCAGAGAGTGTGTTAAGA 300
QY 321 GATGGCAGAACTTCCAAAAGACATTTTGTCTATTTTTCACAGATTCAGAAATCA 380
DB 301 GATGGCAGAACTTCCAAAAGACATTTTGTCTATTTTTCACAGATTCAGAAATCA 360
QY 381 TGTGACATTTTATGACATCTTAAAGTCTGTAAGATATATGTTTCAATTAAGAAAC 440
DB 361 TGTGACATTTTATGACATCTTAAAGTCTGTAAGATATATGTTTCAATTAAGAAAC 420
QY 441 TTATGAAGAACTACAGAGTCAAGATTTGATGTTGTTTTCAGAGTGTGTTTCCCT 500
DB 421 TTATGAAGAACTACAGAGTCAAGATTTGATGTTGTTTTCAGAGTGTGTTTCCCT 480
QY 501 TTGATGAGCTGTGGCCGAGTTACTTAAATACCTTTTGTACAGAGCTCCGCTTCTCTC 560
DB 481 GTATGAGAGCTGTGGCCGAGTTACTTAAATACCTTTTGTACAGAGCTCCGCTTCTCTC 540

QY 561 CTGGCTACGCAATTTGAAAAGCATAGTGGAGACTTGTCTTCTCTTATGTGCTTG 620
DB 541 CTGGCTACACTTTTGAAGAAACATAGTGGAGATTTATTTTCCCTCTCTTACTACT 600
QY 621 TTGTTATGTACAGAACTAAGTGAACCAATGACTTTCATAGAGAGTAAATATGATCT 680
DB 601 TTGTTATGTACAGAACTAAGTGAACCAATGACTTTCATAGAGAGTAAATATGATCT 660
QY 681 ATGTGCTTTATTTTGAATTTTGGTCCAAATATTTGACATGAAGAGTGGATCTCT 740
DB 661 ATGTGCTTTATTTGACTTTTGGTCCAAATATTTGACATGAAGAGTGGATCTCT 720
QY 741 ACACTGAAGTTCTAGAGAAACCACTAGTATCTGACATGCAAAAGCTGACATAT 800
DB 721 ATAGTGAAGTTCTAGAGAAACCACTAGTATCTGACATGCAAAAGCTGACATAT 780
QY 801 GGCCTTATTCGAAACTACTGGGATTTTCAATTTTCTGACCACTTTTCCAAATGTTAGT 860
DB 781 GGCCTTATTCGAAACTACTGGGATTTTCAATTTTCTGACCACTTTTCCAAATGTTAGT 840
QY 861 TCGTTGAGAGACTCCACTGCAAACTGCAAAACCCCTACGGAAGAAATGGAAGTTTG 920
DB 841 TTGTTGAGAGACTCCACTGCAAACTGCAAAACCCCTACGGAAGAAATGGAAGCTTTG 900
QY 921 TCCAGACCTCTGAGAAATAGTGTGTGTGTTTCTCTGGGGTCAATGCTCAGTAACA 980
DB 901 TACAGAGCTCTGAGAAATAGTGTGTGTGTTTCTCTGGGGTCAATGCTCAGTAACA 960
QY 981 CGTGAGAAAGAGGCCCAATGTAATTCATCAAGCCCTTTCGCAATCCCAAAAGGTC 1040
DB 961 TGACAGAAAGAGGCCCAATGTAATTCATCAAGCCCTTTCGCAATCCCAAAAGGTC 1020
QY 1041 TGT-GGAGATTTGAT-GGGAATAACAGATACTTTAGACTCAATCTCGGCTGTACAA 1098
DB 1021 TGTGGAGATTTGATGGGAGATTTAAACAGATACCTTTAGTCTCAATCTCGGCTGTACAA 1080
QY 1099 GT-GGATACCCACAGA-TGATCTTCTGTGTATCCCAAAACCAAGCTTTTATCACTCAT 1156
DB 1081 GTGGAGATACCCACAGAATTTGACTTCTAGTCTCAAAAGACCAAGCTTTTATTAATCAT 1140
QY 1157 GGTGAATGAATGGAGATCTATGAAGCTATTTTCAATGGGGTCCCTATAGTGGAGTTCCC 1216
DB 1141 GGTGAAGCAATGGCACTCTAGAGCAATCTCAATGAGGATCCCTATAGTGGAGATTTCCA 1200
QY 1217 ATATTGGTATCTGAGCTTATTAATAGTCAATGAGGAGGCTTTGAGAACATGATTAAC 1276
DB 1201 TTGTTTGGCATCAACTGATTAATCTGATCAATGAAGGCCAGAGGAGCATGTTAGA 1260
QY 1277 ATAACTTCAAAACTATGACAAAGGAAGTCTTGAAGGCTTTGAGAACATGATTAAC 1336
DB 1261 GTGACCTTCAACCAATGTGAGTACAGACTTGTGATGATGAGAGAGTAAATTAAT 1320
QY 1337 GATTCTCTTTTAAAGAGATGCTATGAGATTTTCAAGAAATTTCAACATGATCAACCTGTA 1396
DB 1321 GATCTCTTATTAAGAAATGTTATGAAATTTTCAAGAAATTTCAACATGATCAACCTGTA 1380
QY 1397 AAGCCCTGATGAGAGAGCTCTTCTGATGAGTGTTCATGCGCCCAAAAGAGCCAAAG 1456
DB 1381 AAGCCCTGATGAGAGAGCTCTTCTGATGAGTGTTCATGCGCCCAAAAGAGCTAA 1440
QY 1457 CACCTGCAATGAGTCCCATGACCTCACTGTTTCCAGACTCTTATGATGATGATTT 1516
DB 1441 CACCTTGGGTTTGAAGCCCAAGACCTCACTGTTTCCAGACTCTTATGATGATGATTT 1500
QY 1517 GGGTTCCTGCTGACCTGTGTGGCAATGCTATATTTCTGTTCAAAAATGTTTTTATTT 1576
DB 1501 GGGTTCCTGCTGCTGTGTGTGGCAATGCTATATTTCTGTTCAAAAATGTTTTGCTGTT 1560
QY 1577 TCTCTCAAAAATTTTAAATTAAGTAAAGATGAGAAAGAGGAAATAGATCTTTG 1631
DB 1561 TCTTCTGGAAGTTTGTGTAAGAAAGCAAAAGGAGAAATGATTAATATATC 1615

RESULT 15
AAS69711
ID AAS69711 standard; cDNA; 2802 BP.
XX AC AAS69711;
XX DT 13-FEB-2002 (first entry)
XX DE DNA encoding novel human diagnostic protein #515.
XX KM Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX KM food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX OS Homo sapiens.
XX PN WO200175067-A2.
XX PD 11-OCT-2001.
XX PF 30-MAR-2001; 2001WO-US08631.
XX PR 31-MAR-2000; 2000US-0540217.
XX PR 23-AUG-2000; 2000US-0649167.
XX PA (HYSE-) HYSEQ INC.
XX PI Dermanac RT, Liu C, Tang YT;
XX DR WPI, 2001-639362/73.
XX DR P-PSDB; ABG05524.
XX PT New isolated polynucleotide and encoded polypeptides, useful in
XX PT diagnostics, forensics, gene mapping, identification of mutations
XX PT responsible for genetic disorders or other traits and to assess
XX PT biodiversity.
XX PS Claim 1; SEQ ID No 5515; 103bp; English.
XX CC The invention relates to isolated polynucleotide (I) and
XX CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
XX CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
XX CC and gene mapping, and in recombinant production of (II). The
XX CC polynucleotides are also used in diagnostics as expressed sequence tags
XX CC for identifying expressed genes. (I) is useful in gene therapy techniques
XX CC to restore normal activity of (II) or to treat disease states involving
XX CC (II). (II) is useful for generating antibodies against it, detecting or
XX CC quantitating a polypeptide in tissue, as molecular weight markers and as
XX CC a food supplement. (II) and its binding partners are useful in medical
XX CC imaging of sites expressing (II). (I) and (II) are useful for treating
XX CC disorders involving aberrant protein expression or biological activity.
XX CC The polypeptide and polynucleotide sequences have applications in
XX CC diagnostics, forensics, gene mapping, identification of mutations
XX CC responsible for genetic disorders or other traits to assess biodiversity
XX CC and to produce other types of data and products dependent on DNA and
XX CC amino acid sequences. AAS64197-AAS94564 represent novel human
XX CC diagnostic coding sequences of the invention.
XX CC Note: The sequence data for this patent did not appear in the printed
XX CC specification, but was obtained in electronic format directly from WIPO
XX CC at ftp.wipo.int/pub/published_pct_sequences.
SQ Sequence 2802 BP; 791 A; 605 C; 571 G; 835 T; 0 other;
Query Match 72.7%; Score 1199.4; DB 23; Length 2802;
Best Local Similarity 85.0%; Pred. No. 0;
Matches 1367; Conservative 0; Mismatches 236; Indels 6; Gaps 2;
QY 26 TGCATCAGATGCTATGAATGAGCTTCACTCTTCCGATACAGCTGAGCTGTAC 85
DB 2 TGCACAGAGATGCTCTGAATGAGCTACAG--TTCTGCTGATACACATCAGTTTTC 58
QY 86 TTAGCTCTGGAGCTTGGAAGAGTGTGTGTGTGGCCACAGAACTTCAAGCCACTGATG 145
DB 59 TTAGCTCTGGAGCTTGGAAGAGTGTGTGTGTGGCCACAGAACTTCAAGCTTGTGATG 118

QY 146 AATATAAGACAATCTCGATGAATCTGTCCAGAGAGTATGAGTATGATGCA 205
DB 119 AATATGAAGACAATCTCGAAGAACTTGTTCAGAGAGTATGAGTATGATGCA 178
QY 206 TCTTCAGCTTCCATTTCTTTCGATCCCAAGCCCATCTCTTAATTTGAAGTTAT 265
DB 179 TCTTCAGCTTCCATTTCTTTCGATCCCAAGCCCATCTCTTAATTTGAAGTTAT 238
QY 266 CCTGTATCTTTAACTAAACTGAGTTTGAGATATTTCAAGCAGCTGTGTAAGATGG 325
DB 239 CCAATATCTTTAACTAAACTGAGTTTGAGATATTTCAAGCAGCTGTGTAAGATGG 298
QY 326 GCAAACTTCCAAAAGACATTTTGGTCAATTTTTCACAAAGACAAAGAAATCATGTGG 385
DB 299 TCAGAAATTTCAAAAAGATACATTTTGGTCAATTTTTCACAAAGACAAAGAAATCATGTGG 358
QY 386 ACAATTAATGACATCTTAGAAAAGTTCTGAAGATATAGTTTCAATTAAGAACTTATG 445
DB 359 GCAATTAATGACATCTTAGAAAAGTTCTGAAGATATAGTTTCAATTAAGAACTTATG 418
QY 446 AAGAACTTACAGAGATCAAGATTTGATGTGTCTTCAAGATGTGTTTCCCTTTGGT 505
DB 419 AAAAACTTACAGAGATCAAGATTTGATGTGTCTTCAAGATGTGTTTCCCTTTGGT 478
QY 506 GAGCTGCTGCGGAGTTACTTAAATACCTTTGTCTACAGCCCTCGCTCTGCTGGC 565
DB 479 GAGCTGCTGCGGAGTTACTTAAATACCTTTGTCTACAGCCCTCGCTCTGCTGGC 538
QY 566 TAGCAATTAAGAAAGCATAGTGAAGAGACTTCTGTTCCCTCTCTTAATGTGCTGTGTT 625
DB 539 TACTCATTTGAAGAGCAGATGAGATTTATTTCCCTCTCTTAATGTGCTGTGTT 598
QY 626 ATGTGAGACTTAAGTGAACCAATGACTTTTATAGAGAGGTAAATAATGATCTATGTG 685
DB 599 ATGTGAGACTTAAGTGAACCAATGACTTTTATAGAGAGGTAAATAATGATCTATGTG 658
QY 686 CTTTATTTTGAATTTTGGTTCCTCAATTAATGATGAAGAAAGTGGATTCAGTTCTACAGT 745
DB 659 CTTTATTTTGAATTTTGGTTCCTCAATTAATGATGAAGAAAGTGGATTCAGTTCTACAGT 718
QY 746 GAAGTTCTAGGAAGACCCCTAGCTTATCTGAGACATGCAAAAGCTGACATATGCTT 805
DB 719 GAAGTTCTAGGAAGACCCCTAGCTTATCTGAGACATGCAAAAGCTGACATATGCTT 778
QY 806 ATTGGAATCTAGTGGAGTTTCAATTTCTCACCCACTTTACCAATGTTGAGTTGCTT 865
DB 779 ATGCGAAATCTCGGAATTTTAATTTCCATCCATCTTACCAATGTTGAGTTGCTT 838
QY 866 GAGGAGCTCCAGTCGCAAACTGCGCAAACTGCGCAAACTGCGCAAACTGCGCAAACTGCGCA 925
DB 839 GAGGAGCTCCAGTCGCAAACTGCGCAAACTGCGCAAACTGCGCAAACTGCGCAAACTGCGCA 898
QY 926 AGCTCTGGAAGAAATGTTGT 985
DB 899 AGCTCTGGAAGAAATGTTGT 958
QY 986 GAAAGAAAGGCGCAATGTAATTTGATGAGCCCTTCCCAAGATCCCAAAAGTTCTGTG- 1044
DB 959 GAAAGAAAGGCGCAATGTAATTTGATGAGCCCTTCCCAAGATCCCAAAAGTTCTGTG- 1018
QY 1045 --GAGATTTGATGAGGAAATTAACAGATACTTTAGAGTCAATCTGCGCTGTCAAGTGG 1102
DB 1019 GAGATTTTGAATGAGGAAATTAACAGATACTTTAGAGTCAATCTGCGCTGTCAAGTGG 1078
QY 1103 ATACCCAGAAATATCTTTGCTCATCCCAAAACCAAAAGCTTTTATCACTACATGATGGA 1162
DB 1079 ATACCCAGAAATATCTTTGCTCATCCCAAAACCAAAAGCTTTTATCACTACATGATGGA 1138
QY 1163 ATGAATGAGATCTATGAAGTATTTACATGAGGATCTTATGATGAGGATTTCCCATATTTT 1222
DB 1139 GCCAATGAGATCTATGAAGGAAATCTACATGAGGATCTTATGATGAGGATTTCCCATATTTT 1198

```
Oy 1223 GGTGATCAGCTTGATTAACATAGTCTCATGTGAAGGCCAAAGAGAGCAGCTGTAGAATAAAC 1282
    |||||
Db 1199 TTGATCAACCTGATTAATATGCTCATGTGAAGGCCAAAGAGAGCAGCTGTAGAGTGAC 1258
    |||||
Oy 1283 TTCAAACTATGACAAGGAGATTACTGAGGGCTTTGAGAACAGTCATTACCGATTCC 1342
    |||||
Db 1259 TTCAAACAATGTGAGTACAGACCTGCTGAATGCACTGAAGACAGTAAATTAATGATCCT 1318
    |||||
Oy 1343 TCTTATAAGGAATGCTATGAGATTATCAAGAAATTCACCATGATCAACCTGTAAAGCCC 1402
    |||||
Db 1319 TCATATATAAGGAATATTATGAAATTAACAAGAAATTCACCAATGATCAACAGTGAAGCCC 1378
    |||||
Oy 1403 CTGATCGAGCAGTCTTCTGATCGAGTTGTCAATGCCCAAAAGAGCCAGCACCTG 1462
    |||||
Db 1379 CTGATCGAGCAGTCTTCTGATCGAGTTGTCAATGCCCAAAAGAGCCAGCACCTT 1438
    |||||
Oy 1463 CGATCAGCTGCCCATGACCTGACCTGTTCCAGACATACTATAGATGATGGGTTG 1522
    |||||
Db 1439 CGAGTTGACGCCCAACACCTCACCCTGGTCCAGTACCACTCTTGATGATGGGTTG 1498
    |||||
Oy 1523 CTGCTGACCTGTGTGGCAACTGCTATATTCTTTGTTCAAAATGTTTTTATTTCCGT 1582
    |||||
Db 1499 CTGCTGCTGTGTGGCAACCGTGTATTATCATAAAGTGTGTGTGTTGTTTTC 1558
    |||||
Oy 1583 CAAAAATTATAAATACTAGAAAGATAGAAAAGAGGGAATAGATCTTTC 1631
    |||||
Db 1559 TGGAACTTGTCTAGAAAAGGAAAGGGAAGGGAATAGTATATC 1607
    |||||
```

Search completed: December 7, 2003, 05:47:47
Job time : 355 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 7, 2003, 05:31:12 ; Search time 90 Seconds

(without alignments)
8092.024 Million cell updates/sec

Title: US-09-980-729B-10

Perfect score: 1650

Sequence: 1 agcaactggaagcaagacat.....ccaattccaagaagaccctg 1650

Scoring table: IDENTITY NUC

Gap 10.0, Gapext 1.0

Searched: 569978 seqs, 220691566 residues

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : Issued Patents NA:*

- 1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
- 2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
- 3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
- 4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
- 5: /cgn2_6/ptodata/2/ina/PTCUTS_COMB.seq:*
- 6: /cgn2_6/ptodata/2/ina/backfile1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|--------|-------------|--------|----|-------------------|
| 1 | 1445.4 | 87.6 | 2092 | 4 | US-09-356-806-7 |
| 2 | 1247.8 | 75.6 | 1854 | 4 | US-09-356-806-39 |
| 3 | 1159.6 | 70.3 | 2107 | 3 | US-09-180-852-1 |
| 4 | 1126.8 | 68.3 | 1976 | 4 | US-09-356-806-112 |
| 5 | 818 | 49.6 | 1413 | 4 | US-09-813-918-1 |
| 6 | 754.8 | 45.7 | 1323 | 4 | US-09-356-806-1 |
| 7 | 610.8 | 37.0 | 1686 | 4 | US-09-356-806-41 |
| 8 | 572.4 | 34.7 | 1001 | 4 | US-09-671-317-403 |
| 9 | 513 | 31.1 | 2312 | 4 | US-09-356-806-114 |
| 10 | 502.2 | 30.4 | 1001 | 4 | US-09-671-317-412 |
| 11 | 327.4 | 19.8 | 2339 | 5 | PCT-US92-00282-2 |
| 12 | 316 | 19.2 | 2336 | 5 | PCT-US92-00282-1 |
| 13 | 269 | 16.3 | 735 | 4 | US-09-305-856B-17 |
| 14 | 245 | 14.8 | 1001 | 4 | US-09-671-317-427 |
| 15 | 242.8 | 14.7 | 1001 | 4 | US-09-671-317-428 |
| 16 | 197 | 11.9 | 1001 | 4 | US-09-671-317-424 |
| 17 | 195.4 | 11.8 | 596 | 4 | US-09-356-806-45 |
| 18 | 193.8 | 11.7 | 978 | 4 | US-09-356-806-118 |
| 19 | 188.4 | 11.4 | 1589 | 4 | US-09-356-806-6 |
| 20 | 186.4 | 11.3 | 1001 | 4 | US-09-671-317-352 |
| 21 | 185.2 | 11.2 | 1001 | 4 | US-09-671-317-353 |
| 22 | 185.2 | 11.2 | 1001 | 4 | US-09-671-317-354 |
| 23 | 182.2 | 11.0 | 1001 | 4 | US-09-671-317-405 |
| 24 | 153.6 | 9.3 | 746 | 4 | US-09-356-806-2 |
| 25 | 153.6 | 9.3 | 1001 | 4 | US-09-671-317-381 |
| 26 | 153.6 | 9.3 | 1001 | 4 | US-09-671-317-382 |
| 27 | 144.8 | 8.8 | 1602 | 4 | US-09-356-806-117 |

| | | | | | | |
|----|-------|-----|------|---|-------------------|--------------------|
| 28 | 144.6 | 8.8 | 1591 | 4 | US-09-356-806-44 | Sequence 44, Appl |
| 29 | 138.6 | 8.4 | 689 | 4 | US-09-356-806-5 | Sequence 5, Appl |
| 30 | 137.4 | 8.3 | 785 | 4 | US-09-356-806-3 | Sequence 3, Appl |
| 31 | 136 | 8.2 | 391 | 4 | US-09-370-838-21 | Sequence 21, Appl |
| 32 | 134.6 | 8.2 | 1001 | 4 | US-09-671-317-351 | Sequence 351, Appl |
| 33 | 134.2 | 8.1 | 1001 | 4 | US-09-671-317-404 | Sequence 404, Appl |
| 34 | 131 | 7.9 | 1340 | 4 | US-09-356-806-42 | Sequence 42, Appl |
| 35 | 124.6 | 7.6 | 1001 | 4 | US-09-671-317-415 | Sequence 415, Appl |
| 36 | 124.6 | 7.6 | 1001 | 4 | US-09-671-317-417 | Sequence 417, Appl |
| 37 | 124.6 | 7.6 | 1021 | 4 | US-09-356-806-115 | Sequence 115, Appl |
| 38 | 124.2 | 7.5 | 1001 | 4 | US-09-671-317-413 | Sequence 413, Appl |
| 39 | 124.2 | 7.5 | 1001 | 4 | US-09-671-317-414 | Sequence 414, Appl |
| 40 | 124.2 | 7.5 | 1001 | 4 | US-09-671-317-416 | Sequence 416, Appl |
| 41 | 124.2 | 7.5 | 1822 | 4 | US-09-356-806-43 | Sequence 43, Appl |
| 42 | 124 | 7.5 | 983 | 4 | US-09-671-317-386 | Sequence 386, Appl |
| 43 | 122.6 | 7.4 | 480 | 4 | US-09-356-806-116 | Sequence 116, Appl |
| 44 | 116.8 | 7.1 | 1001 | 4 | US-09-671-317-360 | Sequence 360, Appl |
| 45 | 102.6 | 6.2 | 1001 | 4 | US-09-671-317-294 | Sequence 294, Appl |

ALIGNMENTS

RESULT 1
US-09-356-806-7
Sequence 7, Application US/09356806
Patent No. 6586175

GENERAL INFORMATION:

APPLICANT: Penny, Laura
APPLICANT: Galvin, Margaret
APPLICANT: Miller, Andrew
APPLICANT: Reidy, Michael
TITLE OF INVENTION: Genotyping Human
TITLE OF INVENTION: UDP-Glucuronosyltransferase 2B4 (UGT2B4), 2B7 (UGT2B7) and
FILE REFERENCE: SEQ-22PRV2
CURRENT APPLICATION NUMBER: US/09/356,806
CURRENT FILING DATE: 1999-07-20
NUMBER OF SEQ. ID NOS: 164
SOFTWARE: PatsSeq for Windows Version 3.0
SEQ ID NO: 7
LENGTH: 2092
TYPE: DNA
ORGANISM: H. sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: (38)...(1621)
US-09-356-806-7

Query Match 87.6%; Score 1445.4; DB 4; Length 2092;

Best Local Similarity 93.3%; Pred. No. 0; Mismatches 106; Indels 3; Gaps 1;

Matches 1522; Conservative 0;

| | | | |
|----|-----|---|-----|
| QY | 1 | AGCAACGGAACCAACAGCATTCATTCAGAGTCTATGAAATGAGACTTCAGCTCT | 60 |
| DB | 4 | AGCAACGGAACCAACAGCATTCATTCAGAGTCTATGAAATGAGACTTCAGCTCT | 63 |
| QY | 61 | TTCTCTGATACAGCTGAGCTGTTACTTACCTCTGGAGATTGTGAAAGGTCTGCTG | 120 |
| DB | 64 | TTCTCTGATACAGCTGAGCTGTTACTTACCTCTGGAGATTGTGAAAGGTCTGCTG | 123 |
| QY | 121 | GCCACAGAAATTCAGCAGCATTCAGATGATTAAGCAATCTGGATGAACCTTCCAGAG | 180 |
| DB | 124 | GCCACAGAAATTCAGCAGCATTCAGATGATTAAGCAATCTGGATGAACCTTCCAGAG | 183 |
| QY | 181 | AGGTCATGAGTGAAGTATTTGGCATCTTCAGCTTCATTTCTTTCATCCCAACAGCC | 240 |
| DB | 184 | AGGTCATGAGTGAAGTATTTGGCATCTTCAGCTTCATTTCTTTCATCCCAACAGCC | 243 |
| QY | 241 | ATCTACTCTTAATTTGAAGTTATCTGTATCTTTAACTGAACTGAGTTTGAGATAT | 300 |
| DB | 244 | ATCTACTCTTAATTTGAAGTTATCTGTATCTTTAACTGAACTGAGTTTGAGATAT | 303 |

```

QY 301 TATCAAGAGCTGGTTAAGATGGGCGAGAACTTCCAAAGACACATTTGGTCATATTT 360
Db 304 TATCAAGAGCTGGTTAAGATGGGCGAGAACTTCCAAAGACACATTTGGTCATATTT 363
QY 361 TTCAACAAGTACAGAAATCATGTGACATTTAATGACATTTAGAAAATTCTGTAAAGA 420
Db 364 TTCAACAAGTACAGAAATCATGTGACATTTAATGACATTTAGAAAATTCTGTAAAGA 423
QY 421 TATAGTTTCAATTAAGAACTTAAGAAAGAACTACAGAGTCAAGATTTGATGTTGTTCT 480
Db 424 TATAGTTTCAATTAAGAACTTAAGAAAGAACTACAGAGTCAAGATTTGATGTTGTTCT 483
QY 481 TGCAGATGCTGTTTCCCTTGTGAGCTGCGCCGAGTTACTTAATAATACCTTTGT 540
Db 484 TGCAGATGCTGTTTCCCTTGTGAGCTGCGCCGAGTTACTTAATAATACCTTTGT 543
QY 541 CTACAGCTCGGCTTCTCTCTGCTAGCGCAATTGAAAAGCAATGTGAGAGACTTCTGTT 600
Db 544 CTACAGCTCGGCTTCTCTCTGCTAGCGCAATTGAAAAGCAATGTGAGAGACTTCTGTT 603
QY 601 CCTCTCTCTATGCTGCTGTTGTTATGTCAGAACTAAGTACCAAAAGCTTTGATAGA 660
Db 604 CCTCTCTCTATGCTGCTGTTGTTATGTCAGAACTAAGTACCAAAAGCTTTGATAGA 663
QY 661 GAGGTAATAAATATGATCTATGCTTATTTGATTTGTTGTTCCAAATATTTGACAT 720
Db 664 GAGGTAATAAATATGATCTATGCTTATTTGATTTGTTGTTCCAAATATTTGACAT 723
QY 721 GAAAGAGTGGATCAGTTCTACAGTAAAGTTCTAGAAAGCCCACTACGTTATCTGAGAC 780
Db 724 GAAAGAGTGGATCAGTTCTACAGTAAAGTTCTAGAAAGCCCACTACGTTATCTGAGAC 783
QY 781 AATGCAAAAGCTGACATATGCTTATTCGAAACTACAGGATTTTCAATTTCCACACC 840
Db 784 AATGCAAAAGCTGACATATGCTTATTCGAAACTACAGGATTTTCAATTTCCACACC 843
QY 841 ACTCTTACCAATATGTTGAGTTGTTGAGAGACTCCACTGCAAACTGCAAACTCCCTAC 900
Db 844 ACTCTTACCAATATGTTGAGTTGTTGAGAGACTCCACTGCAAACTGCAAACTCCCTAC 903
QY 901 GAAAGAAATGGAAGATTGTTCCAGAGCTCTGAGAGAAATGTTGTTGTTGTTCTCT 960
Db 904 GAAAGAAATGGAAGATTGTTCCAGAGCTCTGAGAGAAATGTTGTTGTTGTTCTCT 963
QY 961 GGGGTCGATGTCATTAACAGTCAAGTCAAGAAAGGCCAATGTAATGTCAGCCCTTGC 1020
Db 964 GGGGTCGATGTCATTAACAGTCAAGTCAAGAAAGGCCAATGTAATGTCAGCCCTTGC 1023
QY 1021 CAAGATCCCAAAAGTTCTGTGAGATTTGATGGAATAAACCAGATTACTTTAGACT 1080
Db 1024 CAAGATCCCAAAAGTTCTGTGAGATTTGATGGAATAAACCAGATTACTTTAGACT 1083
QY 1081 CAATACTCGGCTGTACAAAGTGAATCCCAAGATGATCTTTGGTCAACCAAAACAG 1143
Db 1084 CAATACTCGGCTGTACAAAGTGAATCCCAAGATGATCTTTGGTCAACCAAAACAG 1143
QY 1141 AGCTTTTATCACTATGTTGGAATGAATGGAATCTATGAGATTAATTAACATGGGCTCC 1200
Db 1144 AGCTTTTATCACTATGTTGGAATGAATGGAATCTATGAGATTAATTAACATGGGCTCC 1203
QY 1201 TATGTTGGAGTTTCCATATTTGTTGATCAGCTTATTAATGATGTCATGAGAGGCTTT 1320
Db 1204 TATGTTGGAGTTTCCATATTTGTTGATCAGCTTATTAATGATGTCATGAGAGGCTTT 1323
QY 1261 AGGAGCAGCTGTAGAAATTAATCTCAAACTATGACAAAGCGAAGTTTATGAGGCTTT 1320
Db 1264 AGGAGCAGCTGTAGAAATTAATCTCAAACTATGACAAAGCGAAGTTTATGAGGCTTT 1323
QY 1321 GAGAACTGATTAACGATTTCTCTTATTAAGAAAGTATGATGATTTATCAAGAAATCA 1380
Db 1324 GAGAACTGATTAACGATTTCTCTTATTAAGAAAGTATGATGATTTATCAAGAAATCA 1383
QY 1381 CCAATATCAACTGTAAAGCCCTTAGATGAGCAGCTCTTCTGATGAGATTTGTATGCG 1440

```

```

Db 1284 TCATGATCAACAGTGAAGCCCTTGTATCGAGCAGTCTTGTGATGAAATTTGTCATGCG 1443
QY 1441 CCAGAAAGAGCCCAAGACACCTGCGATCAGCTGCCCACTGACCTGCTTCCAGCACTA 1500
Db 1444 CCATAAAGAGCCCAAGACACCTTGGGTTGACGCCACAGCACTCCTGTTCCAGTACCA 1503
QY 1501 CTCTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1560
Db 1504 CTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1563
QY 1561 AAAATGTTTATTTTCTGTCAGAAATTTAATTAATTAAGAAAGATGAAAGAGGGA 1620
Db 1564 AAAATGTTTATTTTCTGTCAGAAATTTAATTAATTAAGAAAGATGAAAGAGGGA 1620
QY 1621 ATAGATCTTC 1631
Db 1621 TTAATTAAGTC 1631

RESULT 2
US-09-356-806-39
; Sequence 39, Application US/09356806
; Patient No. 6586175
; GENERAL INFORMATION:
; APPLICANT: Penny, Laura
; APPLICANT: Galvin, Margaret
; APPLICANT: Miller, Andrew
; APPLICANT: Reidy, Michael
; TITLE OF INVENTION: Genotyping Human
; TITLE OF INVENTION: UDP-glucuronosyltransferase 2B4 (UGT2B4), 2B7 (UGT2B7) and
; FILE REFERENCE: SEQ-22PRV2
; CURRENT FILING DATE: 1999-07-20
; NUMBER OF SEQ ID NOS: 164
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 39
; LENGTH: 1854
; TYPE: DNA
; ORGANISM: H. sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (15) ... (1584)
US-09-356-806-39

Query Match 75.6%; Score 1247.8; DB 4; Length 1854;
Best Local Similarity 85.9%; Pred. No. 0;
Matches 1384; Conservative 0; Mismatches 227; Indels 0; Gaps 0;

QY 21 TGCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 80
Db 1 TGCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 60
QY 81 GTTACTTATGCTGCGAGTTGTGGAAGGTGCTGTGTGTGCGCCACAGAAATTCAGCCACT 140
Db 61 TTTGCTTATGCTGTGGAATTTGTGGAAGGTGCTGTGTGTGCGCCACAGAAATTCAGCCACT 120
QY 141 GGATGATATTAAGCAATCTGTGATGAATCTTGTCCAGAGAGGTGATGATGATGATGAT 200
Db 121 GGATGATATTAAGCAATCTGTGATGAATCTTGTCCAGAGAGGTGATGATGATGATGAT 180
QY 201 TGGATCTTCAAGCTTCCATTTCTTTCATCCCAAGCCCACTTACTTAAATTTGAAG 260
Db 181 TGGATCTTCAAGCTTCCATTTCTTTCATCCCAAGCCCACTTACTTAAATTTGAAG 240
QY 261 TTTATCTGATCTTTTAACTTAACTGATGATGATGATGATGATGATGATGATGATGATGAT 320
Db 241 TTTATCCACATCTTTTAACTTAACTGATGATGATGATGATGATGATGATGATGATGATGAT 300
QY 321 GATGGCGAAGCTTCCAAAGACACATTTGTCATATTTTTCACAGTACAGAAATCA 380
Db 301 GATGGTCAGACCTTCCAAAGACACATTTGTCATATTTTTCACAGTACAGAAATCA 360

```

QY 381 TGTGACATTTATGACATCTTAGAAAGTTCTGTAAGATATAGTTTCAATTAAGAAC 440
DB 361 TGTCAATATTTGGTGAAGTAACTAGAAAGTTCTGTAAGATATAGTTTCAATTAAGAAC 420
QY 441 TTTATGAAGAACTACAGAGTCAAGATTTGATTTGTTTCTGCAAGATGCTGTTTCCCT 500
DB 421 TTTATGAAGAAAGTCAAGAGTCAAGATTTGATTTGATTTGTTTCTGCAAGATGCTGTTTCCCT 480
QY 501 TTTGATGAGCTGCTGCGGAGTACTTAAATACCTTTTGTCTAAGAGCTCGCTTCTCTC 560
DB 481 GTATGAGCTGCTGCGGAGTACTTAAATACCTTTTGTCTAAGAGCTCGCTTCTCTC 540
QY 561 CTGGCTACGCAATTTGAAAGCATAGTGAAGAGCTTCTGTTCCCTCTCTATGCTGCTG 620
DB 541 CTGGCTACGCAATTTGAAAGCATAGTGAAGAGCTTCTGTTCCCTCTCTATGCTGCTG 600
QY 621 TTTGATGTCAGAACTAAGTACCAATGACTTTCTATAGAGAGGTAAATAATGATCT 680
DB 601 TTTGATGTCAGAACTAAGTACCAATGACTTTCTATAGAGAGGTAAATAATGATCT 660
QY 681 ATGCTCTTATTTGAAATTTGTTCCAAATTTTGTGATGATGAAGAGTGGATCACTTCT 740
DB 661 ATGCTCTTATCTTGTGACTTTTGTGCAATAATTTGATGAAGAGTGGATCACTTCT 720
QY 741 ACAGTGAAGTTCTAGAAAGCACTACGTTATCTGAGACAAAGGCAAAAGCTGACATAT 800
DB 721 ATATGAGAGTTCTAGAAAGCACTACGTTATCTGAGACAAAGGCAAAAGCTGACATAT 780
QY 801 GGTCTTATTCGAATCTAGGATTTTCAATTTCTCTCACTCTTCACTCAATGCTGAGT 860
DB 781 GGTCTTATTCGAATCTAGGATTTTCAATTTCTCTCACTCTTCACTCAATGCTGAGT 840
QY 861 TCGTTGAGAGCTCACTGCAAACTGCAAACTGCAAACTGCAAACTGCAAACTGCAAA 920
DB 841 TCGTTGAGAGCTCACTGCAAACTGCAAACTGCAAACTGCAAACTGCAAACTGCAAA 900
QY 921 TCGAGAGCTGAGAAATGATGTTGTTGTTTCTGTTGTTGTTGTTGTTGTTGTTGTTG 980
DB 901 TACAGAGCTGAGAAATGATGTTGTTGTTTCTGTTGTTGTTGTTGTTGTTGTTGTTG 960
QY 981 CGTCAGAGAAAGGCAATGATTTGATGATGAGGCTTGTGCAAGATCCCAAAAGGTTT 1040
DB 961 TGAAGAGAAAGGCAATGATTTGATGATGAGGCTTGTGCAAGATCCCAAAAGGTTT 1020
QY 1041 TGTGAGATTTGATGAGATTAACCAAGATCTTGAAGCTCAATGCTGCTGTTCAAGT 1100
DB 1021 TGTGAGATTTGATGAGATTAACCAAGATCTTGAAGCTCAATGCTGCTGTTCAAGT 1080
QY 1101 GGAATACCCCAAGATGATCTTGTGTCATCCCAAAAGGCTTTTATCACTGATGTTG 1160
DB 1081 GGAATACCCCAAGATGATCTTGTGTCATCCCAAAAGGCTTTTATCACTGATGTTG 1140
QY 1161 GAATGAATGGATCTATGAAGCTATTTACATGAGGCTTCCATGATGAGGATTTCCCAT 1220
DB 1141 GAGGCAATGGATCTATGAAGCTATTTACATGAGGCTTCCATGATGAGGATTTCCCAT 1200
QY 1221 TTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1280
DB 1201 TTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1260
QY 1281 ACTTCAAAATCTATGACAAAGCAAGATTTTATGAGAGGCTTTGAGAGCTGTTAGAA 1340
DB 1261 ACTTCAAAATCTATGACAAAGCAAGATTTTATGAGAGGCTTTGAGAGCTGTTAGAA 1320
QY 1341 CCTCTTAAAGAGATGCTATGATGATGATGATGATGATGATGATGATGATGATGATG 1400
DB 1321 CCTCTTAAAGAGATGCTATGATGATGATGATGATGATGATGATGATGATGATGATG 1380
QY 1401 CCTCTTAAAGAGATGCTATGATGATGATGATGATGATGATGATGATGATGATGATG 1460
DB 1381 CCTCTTAAAGAGATGCTATGATGATGATGATGATGATGATGATGATGATGATGATG 1440

QY 1461 TGGATGAGCTGCGCATGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1520
DB 1441 TGGATGAGCTGCGCATGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1500
QY 1521 TCTGCTGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1580
DB 1501 TCTGCTGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1560
QY 1581 GTCAAAATTTATTAAGATGAGAAAGTAAAGATGAGAAAGTAAAGATGAGTCTTC 1631
DB 1561 TCTGAGAGTTGCTGAGAAAGCAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1611

RESULT 3
US-09-180-852-1
Sequence 1, Application US/09180852
Patent No. 6287834
GENERAL INFORMATION:
APPLICANT: BELANGER, Alain
APPLICANT: HUM, Dean W.
APPLICANT: BEAULIEU, Martin
TITLE OF INVENTION: CHARACTERIZATION AND USE OF AN ISOLATED URIDINE
FILE REFERENCE: 1259-449
CURRENT APPLICATION NUMBER: US/09/180,852
CURRENT FILING DATE: 1999-02-08
EARLIER APPLICATION NUMBER: PCT/CA97/00328
EARLIER FILING DATE: 1997-05-16
EARLIER APPLICATION NUMBER: US 08/649,319
NUMBER OF SEQ ID NOS: 2
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 1
LENGTH: 2107
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: (52)..(1644)
US-09-180-852-1

Query Match 70.3%; Score 1159.6; DB 3; Length 2107;
Best Local Similarity 82.2%; Pred. No. 0;
Matches 1346; Conservative 0; Mismatches 289; Indels 3; Gaps 1;
QY 1 AGCAACTGGAAGAAAGCAAGATGATGATGATGATGATGATGATGATGATGATGATGAT 60
DB 18 AACCACTGGAAGAAAGCAAGATGATGATGATGATGATGATGATGATGATGATGATGAT 77
QY 61 TCTCTGATACAGTCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 120
DB 78 TCTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 137
QY 121 GCCCAGAGATTCAGGCACTGATGATGATGATGATGATGATGATGATGATGATGATGAT 180
DB 138 GCCCAGAGATTCAGGCACTGATGATGATGATGATGATGATGATGATGATGATGATGAT 197
QY 181 AGGTCATGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 240
DB 198 GGGTCATGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 257
QY 241 ATCTACTCTTAAATTTGAAAGTTATCTGATCTTTTAACTGAAAGTTGAGATAT 300
DB 258 ATCTGCTATTTAAATTTGAAAGTTATCTGATCTTTTAACTGAAAGTTGAGATAT 317
QY 301 TATCAAGCAGCTGTTAAGAGATGAGGAGAG---ACTTCAAAAGACATTTTGGTCATA 357
DB 318 TTTTATGAAGATGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 377
QY 358 TTTTTCAGAGTACAGAAATTCATGATGATGATGATGATGATGATGATGATGATGATGAT 417
DB 378 TTTTTCAGAGTACAGAAATTCATGATGATGATGATGATGATGATGATGATGATGATGAT 437

```
Oy 418 GGATATAGTTTCAAAATAGAACTTATGAGAACTACAGAGTCAAGATTGATGTTGT 477
Db 438 AGATGACGTTTGAACAAAGAACTTATGAGAACTACAGAGTCAAAATTTGATGTCT 497
Oy 478 TCTTGACAGTGTGTTTTCCCTTGTGAGCTGTGCGCGAGTACTTAAATACCTT 537
Db 498 TCTGACAGTGTGTTTTCCCTTGTGAGCTGTGCGCGAGTACTTAAATACCTT 557
Oy 538 TGTCTACAGCTGTGTTTTCCCTTGTGAGCTGTGCGCGAGTACTTAAATACCTT 597
Db 558 TCTGACAGTGTGTTTTCCCTTGTGAGCTGTGCGCGAGTACTTAAATACCTT 617
Oy 598 GTTCCCTCTTCTCTGAGCTGTGAGCTGTGAGCTGTGAGCTGTGAGCTGTGAG 657
Db 618 GTTCCCTCTTCTCTGAGCTGTGAGCTGTGAGCTGTGAGCTGTGAGCTGTGAG 677
Oy 658 AGAGAGGTAAATAATATGATCTATGCTTATTTGAAATTTGGTTCCAAATATTTGA 717
Db 678 GAGAGAGTAAATAATATGATCTATGCTTATTTGAAATTTGGTTCCAAATATTTGA 737
Oy 718 CATGAGAGTGTGAGTCTAGTCTACAGTGAAGTTCTAGAGAGCCCTACCTATTTGA 777
Db 738 TCTGAGAGTGTGAGTCTAGTCTACAGTGAAGTTCTAGAGAGCCCTACCTATTTGA 797
Oy 778 GACAAATGGCAAAAGCTGACATATGCTTATTTGAAATTTGGTTCCAAATATTTGA 837
Db 798 GACAAATGGCAAAAGCTGACATATGCTTATTTGAAATTTGGTTCCAAATATTTGA 857
Oy 838 CCCACTCTTACCAATGTTGAGTGTGAGAGCTGACATGCAAACTGCAAACTGCA 897
Db 858 CCCACTCTTACCAATGTTGAGTGTGAGAGCTGACATGCAAACTGCAAACTGCA 917
Oy 898 ACCGAGAGAAATGAGAGTGTGAGAGCTGAGAGCTGAGAGAGAGAGAGAGAGAG 957
Db 918 GCTTAAAGAGAAAGAGAGTGTGAGAGCTGAGAGAGAGAGAGAGAGAGAGAGAG 977
Oy 958 TCTGGGCTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 1017
Db 978 TCTGGGCTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 1037
Oy 1018 TGCCAAAGATCCCAAAAGTTCTGTGAGATTTGATGGAATTAACAGATACTTTAG 1077
Db 1038 TGCCAAAGATCCCAAAAGTTCTGTGAGATTTGATGGAATTAACAGATACTTTAG 1097
Oy 1078 ACTCAATCTGCGCTGTACAGTGTATCCCAAGATGATCTTCTTGGTCAATCCAAAC 1137
Db 1098 TTCCAAATCTGCGCTGTACAGTGTATCCCAAGATGATCTTCTTGGTCAATCCAAAC 1157
Oy 1138 CAAGCTTTTATCACTCATGTGAGTGAATGAGTGTGAGTGTGAGTGTGAGTGTGAG 1197
Db 1158 CAAGCTTTTATCACTCATGTGAGTGAATGAGTGTGAGTGTGAGTGTGAGTGTGAG 1217
Oy 1198 CCTTATGTTGGAGTGTCCCATATTTTGTGATCAGCTTGTATTAATAGTCATGAGAG 1257
Db 1218 CCTTATGTTGGAGTGTCCCATATTTTGTGATCAGCTTGTATTAATAGTCATGAGAG 1277
Oy 1258 CAAGAGAGAGCTGTGAGAAATTAATCTTCAAACTTGTACAGAGAGAGATTTACTGAGGC 1317
Db 1278 CAAGAGAGAGCTGTGAGAAATTAATCTTCAAACTTGTGTCAAGAGAGATTTGCTCAATGC 1337
Oy 1318 TTTGAGAACAGTCAATTAACGATTTCTCTTATTAAGAGATGCTATGAGATTAACAAGAT 1377
Db 1338 ATTTGAGAACAGTCAATTAACGATTTCTCTTATTAAGAGATTAACAAGATTAACAAGAT 1397
Oy 1378 TCACCATGATCAACCTGTAAAGCCCTAGATGAGCAGATCTTCTGATGAGATTTGTGAT 1437
Db 1398 TCATATGATCAACCGGTGAAGCCCTGATGAGCAGATCTTCTGATGAGATTTGTGAT 1457
Oy 1438 GGGCCACAAAGAGCCAAAGCCTGTGATCAGCTGCCATGACCTTCACTGTTCCAGCA 1497
Db 1458 GGGCCATAAAGAGCCAAAGCCTTCTGATGAGCAGCCCAACCTTCACTGATCAGTA 1517
```

```
Oy 1498 CTACTATAGATGATGATGGGTTCTGTGAGCTGTGAGCACTGTATATTCTTGT 1557
Db 1518 CCACTTTTGGATGTGATGATCTTCTGTGAGCTGTGAGCACTGTATATTATGAT 1577
Oy 1558 CACAAATGTTTTTATTTTCTGTCAAAATTTAATTAAGTGAAGAGAGAG 1617
Db 1578 CACAAATGTTGCTGTGTTGTTTCCGAAAGCTTCCGAAAGAGAGAGAGAG 1637
Oy 1618 GGAATAGATCTTCCAA 1635
Db 1638 GGAATAGATTAATCAAA 1655

RESULT 4
US-09-356-806-112
; Sequence 112: Application US/09356806
; Patent No. 6586175
; GENERAL INFORMATION:
; APPLICANT: Penny, Laura
; APPLICANT: Galvin, Margaret
; APPLICANT: Miller, Andrew
; APPLICANT: Reidy, Michael
; TITLE OF INVENTION: Genotyping Human
; TITLE OF INVENTION: UDP-Glucuronosyltransferase 2B4 (UGT2B4), 2B7 (UGT2B7) and
; TITLE OF INVENTION: 2B15 (UGT2B15) Genes
; FILE REFERENCE: SEQ-22PRV2
; CURRENT APPLICATION NUMBER: US/09/356,806
; CURRENT FILING DATE: 1999-07-20
; NUMBER OF SEQ ID NOS: 164
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 112
; LENGTH: 1976
; TYPE: DNA
; ORGANISM: H. sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (11) ... (1598)
US-09-356-806-112

Query Match 68.3%; Score 1126.8; DB 4; Length 1976;
Best Local Similarity 81.7%; Pred. No. 0;
Matches 1315; Conservative 0; Mismatches 292; Indels 3; Gaps 1;

Oy 29 ATCAGATGTGTATGAATGATGATGATGATGATGATGATGATGATGATGATGAT 88
Db 5 ACCAGATGTGTGTGAATGATGATGATGATGATGATGATGATGATGATGATGAT 64
Oy 89 AGCTGTGAGTGTGAGAGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 148
Db 65 AGCTGTGAGTGTGAGAGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 124
Oy 149 ATTAAGACATCTGTGATGATGATGATGATGATGATGATGATGATGATGATGAT 208
Db 125 ATTAAGACATCTGTGATGATGATGATGATGATGATGATGATGATGATGATGAT 184
Oy 209 TCAGCTTCATTTCTTTCATGATCCCAAGAGCCATCTCTTAAATTTGAAGTTATCT 268
Db 185 TCGGCTTCTACTTGTTCATGATCCCAAGAGCCATCTCTTAAATTTGAAGTTATCT 244
Oy 269 GTATCTTTAACTAAATCTGATGATGATGATGATGATGATGATGATGATGATGAT 328
Db 245 ACATCTTTAACTAAATCTGATGATGATGATGATGATGATGATGATGATGATGAT 304
Oy 329 ---GACTTCCAAAGACATCTTGTGATGATGATGATGATGATGATGATGATGATG 385
Db 305 TATGATGTTTCAAAATTAATGATGATGATGATGATGATGATGATGATGATGATG 364
Oy 386 ACATTTATGACATCTTGAAGAGTGTGATGATGATGATGATGATGATGATGATGAT 445
Db 365 GATATTTATGACATCTTGAAGAGTGTGATGATGATGATGATGATGATGATGATGAT 424
Oy 446 AAGAACTACAGAGATCAAGATTTGATGATGATGATGATGATGATGATGATGATG 505
```


Db 425 ATGAACTACAGAGTCAGAACTTGTGATCTCTGCGAGATGCCCTTAATCCCTGAGT 484
 Qy 506 GAGCTGCTGCGCGAGTTACTTAATAATACCTTGTCTACAGCTCCGCTCTCTCCGGC 565
 Db 485 GAGCTACTGCTGACATTTTAACTACACCTCTTCTGACAGCTCTGATCTCTGTTGGC 544
 Qy 566 TACGCAATGAAAAGCATAGTGAAGAGCTTCTGTTCCCTCCCTTCTATGCTCTGTTGT 625
 Db 545 TACACATTTGAGAGAAATGATGAGAGATTTCTGTTCCCTCTCTATGATCCTGTTGT 604
 Qy 626 ATGTCAGACTAATGATGACCAATGATCTTCAATAGAGGGTAAATAATATGATCTATGTG 685
 Db 605 ATGTCAGAAATTAATGATCAATGATTTTCAATGAGAGGATAAATAATATATACATATG 664
 Qy 686 CTATATTTGAAATTTGGTCCAAATATTTGACATGAAAGATGGAGATCACTTCTACAGT 745
 Db 665 CTATATTTGACTTTGGTTCCTTCAAAATTTATATATCTGAAAGATGGAGCACTTTTATAGT 724
 Qy 746 GAACTTCTAGAGAAACCCACTACGTTATCTGAGCAATGCGAAAGCTGACATATGCTT 805
 Db 725 GAACTTCTAGAGAAACCCACTACATATTTGAGCAATGCGAAAGCTGAAATGTTGCTC 784
 Qy 806 ATTCGAAACTACAGGATTTTCAATTTCTGACCCACTCTTACCAATGTTGAGTCTGT 865
 Db 785 ATTCGAACTATTTGGGATTTTGAATTTCTGCGCCACTCTTACCAAAATGTTGATTTGTT 844
 Qy 866 GGAGAGCTCCCTGCGCAAACTGCGCAAACTCCCTACCGAAGAAATGGAAGATTTGTCCAG 925
 Db 845 GGAGAGCTCCCTGCGCAAACTGCGCAAACTCCCTGCGCAAAAGGAAAGATTTGTGAG 904
 Qy 926 AGCTCTGAGAGAAATGCTGTTGTGTTCTCTGCGGGTGCATGATGATGATGATGATGATG 985
 Db 905 AGCTCTGAGAGAAATGCTGTTGTGTTCTCTGCGGGTGCATGATGATGATGATGATGATG 964
 Qy 986 GAAAGAAAGGCGCAATGATTTGATGATGATGATGATGATGATGATGATGATGATGATG 1045
 Db 965 GAAAGAAAGTGCACATGATTTGATGATGATGATGATGATGATGATGATGATGATGATG 1024
 Qy 1046 AGATTTGATGGGAATTAACCGATATCTTGAAGCTCAATATCTGCGCTGTATCAAGTGATA 1105
 Db 1025 AGATTTGATGGGAATTAACCGATATCTTGAAGCTCAATATCTGCGCTGTATCAAGTGATA 1084
 Qy 1106 CCCGAGATGATCTTCTGCTGATCCCAAAACCAAAAGCTTTATCACTCATGATGATGATG 1165
 Db 1085 CCCGAGATGATCTTCTGCTGATCCCAAAACCAAAAGCTTTATCACTCATGATGATGATG 1144
 Qy 1166 AATGGATCTATGAGATTAATTAACATGAGGATGCTTATGATGATGATGATGATGATG 1225
 Db 1145 AATGGATCTATGAGGATTAATTAACATGAGGATGCTTATGATGATGATGATGATGATG 1204
 Qy 1226 GATGAGCTTATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1285
 Db 1205 GATGAGCTTATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1264
 Qy 1286 AAAACATGACAGCAAGATTTAATGAGGCTTTGAGAAAGCTCAATTAACGATTTCCCT 1345
 Db 1265 AGCAACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1324
 Qy 1346 TATTAAGAGATGCTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1405
 Db 1325 TATTAAGAGATGCTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1384
 Qy 1406 GATGAGCACTTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1465
 Db 1385 GATGAGCACTTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1444
 Qy 1466 TCAGCTGCCATGACCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1525
 Db 1445 GTCGCACTGACCAACCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1504
 Qy 1526 CTGACCTGTGTGAGCACTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1585
 Db 1505 CTGACCTGTGTGAGCACTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1564

Qy 1586 AAATTAATAAACTAGAAATGAAAGAGGGAATAGATCTTCCAAA 1635
 Db 1565 AAGCTCCAAAACAGAAAGAGAAAGAAAGAGATTGATTATCAAAA 1614
 RESULT 5
 US-09-813-918-1
 : Sequence 1, Application US/09813918
 : Patent No. 6383789
 : GENERAL INFORMATION:
 : APPLICANT: WEBSTER, Marion et al.
 : TITLE OF INVENTION: ISOLATED HUMAN DRUG-METABOLIZING
 : TITLE OF INVENTION: PROTEINS, NUCLEIC ACID MOLECULES ENCODING HUMAN
 : TITLE OF INVENTION: DRUG-METABOLIZING PROTEINS,
 : TITLE OF INVENTION: AND USES THEREOF
 : FILE REFERENCE: CL001175
 : CURRENT APPLICATION NUMBER: US/09/813,918
 : CURRENT FILING DATE: 2001-03-22
 : NUMBER OF SEQ ID NOS: 4
 : SOFTWARE: FASTSEQ for Windows Version 4.0
 : SEQ ID NO 1
 : LENGTH: 1413
 : TYPE: DNA
 : ORGANISM: Human
 US-09-813-918-1

Query Match 49.6%; Score 818; DB 4; Length 1413;
 Best Local Similarity 72.9%; Pred. No. 3e-229;
 Matches 1182; Conservative 0; Mismatches 215; Indels 225; Gaps 1;

Qy 13 ACAAGCATTCATTCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 72
 Db 1 AGAATGATCAATTCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 60
 Qy 73 GCTGAGCTGTACTTACTTACTGCTCTGAGAGTGTGAAAGTCTGATGATGATGATGATGATG 132
 Db 61 TCTCAGTTGTACTTACTTACTGCTCTGAGAGTGTGAAAGTCTGATGATGATGATGATGATG 120
 Qy 133 CAGCAGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 192
 Db 121 CAGCAGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 180
 Qy 193 GACTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 252
 Db 181 GACTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 240
 Qy 253 ATTTGAAGTTATCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 312
 Db 241 ATTTGAAGTTATCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 300
 Qy 313 GGTTAAGATGAGGAGCAAGCTTCCAAAAGACATTTTGTGATATTTTTCACAAAGTACA 372
 Db 301 GGTTAAGATGAGGAGCAAGCTTCCAAAAGACATTTTGTGATATTTTTCACAAAGTACA 360
 Qy 373 AGAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 432
 Db 361 AGAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 420
 Qy 433 TAAGAACTATGAGAAAGATGACAGAGTCAAGATTTGATGATGATGATGATGATGATGATGATG 492
 Db 421 TAAGAACTATGAGAAAGATGACAGAGTCAAGATTTGATGATGATGATGATGATGATGATGATG 480
 Qy 493 TTTTCCCTTTGTGAGCTGTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTG 552
 Db 481 TTTTCCCTGTGAGCTGTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAG 521
 Qy 553 CTTCCTCTGCTGAGCAATTTGAAAAAGCATAGTGAAGACTTCTGTTCCCTCTCTCTA 612
 Db 522 ----- 521
 Qy 613 TGTGCTGTGTTATGTCAGAACTAAGTACCAAAATGATCTTATAGAGAGGTAATAAAA 672

```

Db 522 ----- 521
Qy 673 TATGATCTATGCTTATTTTGAATTTTGGTCCAAATATTTGACATGAAGAGTGGGA 732
Db 522 ----- 521
Qy 733 TCAGTCTACAGTGAAGTTCTAGAGAGACCCACTACGTTATCTGAGCAATGGCAAAAGC 792
Db 522 ----- 521
Qy 793 TCAGATATGCTTATTCGAAACTACTGGGATTTTCAATTTCTCAACCCACTTACCAA 852
Db 522 ----- 521
Qy 793 TCAGATATGCTTATTCGAAACTACTGGGATTTTCAATTTCTCAACCCACTTACCAA 852
Db 522 ----- 521
Qy 556 TGACATATGCTTATTCGAAACTACTGGGATTTTCAATTTCTCAACCCACTTACCAA 615
Db 556 TGACATATGCTTATTCGAAACTACTGGGATTTTCAATTTCTCAACCCACTTACCAA 615
Qy 853 TGTGAGTCTGTTGAGAGACTCCACTGCAACCTGCGCAACCCCTACCGCAAGGAATGGA 912
Db 853 TGTGAGTCTGTTGAGAGACTCCACTGCAACCTGCGCAACCCCTACCGCAAGGAATGGA 912
Qy 616 CGTTGATTTTGTGAGAGATTCCTACTGCAACCTGCGCAACCCCTACCGCAAGGAATGGA 675
Db 616 CGTTGATTTTGTGAGAGATTCCTACTGCAACCTGCGCAACCCCTACCGCAAGGAATGGA 675
Qy 913 AGAGTTTGTCCAGAGCTCTGAGAGAAATGATGTTGTTGTTTCTCTGAGGTCGATGAT 972
Db 913 AGAGTTTGTCCAGAGCTCTGAGAGAAATGATGTTGTTGTTTCTCTGAGGTCGATGAT 972
Qy 676 GGAATTTTACAGAGCTCTGAGAGAAATGATGTTGTTGTTTCTCTGAGGTCGATGAT 735
Db 676 GGAATTTTACAGAGCTCTGAGAGAAATGATGTTGTTGTTTCTCTGAGGTCGATGAT 735
Qy 973 CAGTAAACAGTCAAGAGAGAGGCAATGTAATTCATCAGCCCTTGCCAAAGTCCACA 1032
Db 973 CAGTAAACAGTCAAGAGAGAGGCAATGTAATTCATCAGCCCTTGCCAAAGTCCACA 1032
Qy 736 AAGTAAACAGTCAAGAGAGAGGCAATGTAATTCATCAGCCCTTGCCAAAGTCCACA 795
Db 736 AAGTAAACAGTCAAGAGAGAGGCAATGTAATTCATCAGCCCTTGCCAAAGTCCACA 795
Qy 1033 AAGGTTCTGAGAGATTTGATGGAATTAACCATGATCTTGAAGCTCAATACCTGGCT 1092
Db 1033 AAGGTTCTGAGAGATTTGATGGAATTAACCATGATCTTGAAGCTCAATACCTGGCT 1092
Qy 796 AAGGTTCTGAGAGATTTGATGGAATTAACCATGATCTTGAAGCTCAATACCTGGCT 855
Db 796 AAGGTTCTGAGAGATTTGATGGAATTAACCATGATCTTGAAGCTCAATACCTGGCT 855
Qy 1093 GTACAGATGATATACCCAGAAATGATCTTGGTCAATCCCAAAACCAAGCTTTTATAC 1152
Db 1093 GTACAGATGATATACCCAGAAATGATCTTGGTCAATCCCAAAACCAAGCTTTTATAC 1152
Qy 856 GTACAGATGATATACCCAGAAATGATCTTGGTCAATCCCAAAACCAAGCTTTTATAC 915
Db 856 GTACAGATGATATACCCAGAAATGATCTTGGTCAATCCCAAAACCAAGCTTTTATAC 915
Qy 1153 TCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1212
Db 1153 TCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1212
Qy 916 TCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 975
Db 916 TCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 975
Qy 1213 TCCCATTTTGGTATCAGCTTGAATCAATGATCAATGAGGCAAGGCAAGGCAAGCT 1272
Db 1213 TCCCATTTTGGTATCAGCTTGAATCAATGATCAATGAGGCAAGGCAAGGCAAGCT 1272
Qy 976 TCCATTTGTTTGTATCAACCTGATACATGATGATGATGATGATGATGATGATGAT 1035
Db 976 TCCATTTGTTTGTATCAACCTGATACATGATGATGATGATGATGATGATGATGAT 1035
Qy 1273 AGAATTAACCTCAAACTATGACAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGCT 1332
Db 1273 AGAATTAACCTCAAACTATGACAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGCT 1332
Qy 1036 TAGATTTGATCTTCAACATATGATGATGATGATGATGATGATGATGATGATGAT 1095
Db 1036 TAGATTTGATCTTCAACATATGATGATGATGATGATGATGATGATGATGATGAT 1095
Qy 1333 TACCGATTCCTCTTATTAAGAGAAATGATGATGATGATGATGATGATGATGATGAT 1392
Db 1333 TACCGATTCCTCTTATTAAGAGAAATGATGATGATGATGATGATGATGATGATGAT 1392
Qy 1096 TAAATGATCCTTATTAAGAGAAATGATGATGATGATGATGATGATGATGATGATGAT 1155
Db 1096 TAAATGATCCTTATTAAGAGAAATGATGATGATGATGATGATGATGATGATGATGAT 1155
Qy 1393 TGTAAAGCCCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1452
Db 1393 TGTAAAGCCCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1452
Qy 1156 AGTAAAGCCCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1215
Db 1156 AGTAAAGCCCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1215
Qy 1453 CAAGACCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1512
Db 1453 CAAGACCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1512
Qy 1216 CAAGACCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1275
Db 1216 CAAGACCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1275
Qy 1513 GATTTGGTCTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1572
Db 1513 GATTTGGTCTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1572
Qy 1276 GATTTGGTCTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1335
Db 1276 GATTTGGTCTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1335
Qy 1573 ATTTTCTGCTCAAAATTTAATTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1632
Db 1573 ATTTTCTGCTCAAAATTTAATTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1632
Qy 1336 GTTTTGTCTGAGAGTTGCTAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1395
Db 1336 GTTTTGTCTGAGAGTTGCTAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1395
Qy 1633 AA 1634
Db 1396 GA 1397

```

```

Sequence 1, Application us/09356806
Patent No. 6586175
GENERAL INFORMATION:
APPLICANT: Penny, Laura
APPLICANT: Galvin, Margaret
APPLICANT: Miller, Andrew
APPLICANT: Reidy, Michael
TITLE OF INVENTION: Genotyping Human
TITLE OF INVENTION: UDP-Glucuronosyltransferase 2B4 (UGT2B4), 2B7 (UGT2B7) and
TITLE OF INVENTION: 2B15 (UGT2B15) Genes
FILE REFERENCE: SEQ-22PRV2
CURRENT APPLICATION NUMBER: US/09/356,806
NUMBER OF FILING DATE: 1999-07-20
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO: 1
LENGTH: 1323
TYPE: DNA
ORGANISM: H. sapiens
FEATURE:
NAME/KEY: Other
LOCATION: (140)...(897)
US-09-356-806-1

Query Match 45.7%; Score 754.8; DB 4; Length 1323;
Best Local Similarity 99.7%; Pred. No. 9e-211;
Matches 756; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 AGCACTGGAAAAACAGATTTGATGATGATGATGATGATGATGATGATGATGATGATGAT 60
Db 143 AGCACTGGAAAAACAGATTTGATGATGATGATGATGATGATGATGATGATGATGATGAT 202
Qy 61 TCTCTGATACAGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 120
Db 203 TCTCTGATACAGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 262
Qy 121 GCCCAGAAATTCAGCACTGATGATGATGATGATGATGATGATGATGATGATGATGAT 180
Db 263 GCCCAGAAATTCAGCACTGATGATGATGATGATGATGATGATGATGATGATGATGAT 322
Qy 181 AGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 240
Db 323 AGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 382
Qy 241 ATCTACTCTTAAATTTGAAGTTATCTGATGATGATGATGATGATGATGATGATGATGAT 300
Db 383 ATCTACTCTTAAATTTGAAGTTATCTGATGATGATGATGATGATGATGATGATGATGAT 442
Qy 301 TATCAAGAGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 360
Db 443 TATCAAGAGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 502
Qy 361 TTCACAGTACAGAAATCATGATGATGATGATGATGATGATGATGATGATGATGATGAT 420
Db 503 TTCACAGTACAGAAATCATGATGATGATGATGATGATGATGATGATGATGATGATGAT 562
Qy 421 TATGTTTCAATTAAGAACTATGAGAGAACTACAGAGTCAAGATTTGATGATGATGAT 480
Db 563 TATGTTTCAATTAAGAACTATGAGAGAACTACAGAGTCAAGATTTGATGATGATGAT 622
Qy 481 TGCAGATGCTTTTCCCTTTGGTGAAGCTGCGCCAGATTAATTAATTAATTAATTAAT 540
Db 623 TGCAGATGCTTTTCCCTTTGGTGAAGCTGCGCCAGATTAATTAATTAATTAATTAAT 682
Qy 541 CTACAGCTCGCTTCTCTCTGCTACGCAATGAAAGATGATGAGAGATTTCTGTT 600
Db 683 CTACAGCTCGCTTCTCTCTGCTACGCAATGAAAGATGATGAGAGATTTCTGTT 742
Qy 601 CCTCTCTCTATGAGCTGCTGTTGATGATGATGATGATGATGATGATGATGATGATGAT 660
Db 743 CCTCTCTCTATGAGCTGCTGTTGATGATGATGATGATGATGATGATGATGATGATGAT 802
Qy 661 GAGGTAATAAATATGATCTATGCTTATTTGATGATGATGATGATGATGATGATGAT 720

```

```
Db 803 GAGGTAATAATGATCTATGCTTATTTGAAATTTGTTCCAAATATTGACAT 862
Qy 721 GAAGAAGTGGATAGTTCTTCAGTGAAGTTCTTGAA 758
Db 863 GAAGAAGTGGATAGTTCTTCAGTGAAGTTCTTGAA 900

RESULT 7
US-09-356-806-41
; Sequence 41, Application US/09356806
; Patent No. 6586175
; GENERAL INFORMATION:
; APPLICANT: Penny, Laura
; APPLICANT: Galvin, Margaret
; APPLICANT: Miller, Andrew
; APPLICANT: Reidy, Michael
; TITLE OF INVENTION: Genotyping Human
; TITLE OF INVENTION: UDP-Glucuronosyltransferase 2B4 (UGT2B4), 2B7 (UGT2B7) and
; FILE REFERENCE: SFO-22PRV2
; CURRENT FILING DATE: 1999-07-20
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 41
; LENGTH: 1686
; TYPE: DNA
; ORGANISM: H. sapiens
; FEATURE:
; NAME/KEY: exon
; LOCATION: (392)...(1126)
US-09-356-806-41
```

```
Query Match 37.0%; Score 610.8; DB 4; Length 1686;
Best Local Similarity 87.9%; Pred. No. 1.3e-168;
Matches 666; Conservative 0; Mismatches 92; Indels 0; Gaps 0;

Qy 1 AGCACTCGAAAAAAGCATTCATTCATGATGATGCTCTATGAAATGACCTTACGCT 60
Db 372 AGCAACTCGAAAAAAGCATTCATTCATGATGATGCTCTATGAAATGACCTTACGCT 431
Qy 61 TCTCTGATACAGCTGAGCTGTTACTTATGCTTGGAGTGGAGAAAGTGTGTG 120
Db 432 TTGCTTAATACAGCTGAGCTTCTTATGCTTGGAGTGGAGAAAGTGTGTG 491
Qy 121 GCCCAGCAATTCAGCCATGATGATATATTAAGACATCTCGATGATGATGATG 180
Db 492 GGCAGCAGAAATACAGCCATGATGATATATTAAGACATCTCGATGATGATGATG 551
Qy 181 AGGTGATGAGTGTGCTGATTTGGCATCTTCAGCTTCCATTTCTTCGATCCCAAGCCC 240
Db 552 AGGTGATGAGTGTGCTGATTTGGCATCTTCAGCTTCCATTTCTTCGATCCCAAGCCC 611
Qy 241 ATCTACTCTTAATTTGAGTTTATCCGATCTCTTAACTTAACTGATGATGATGAT 300
Db 612 ATCCGCTCTTAATTTGAGTTTATCCGATCTCTTAACTTAACTGATGATGATGAT 671
Qy 301 TATCAAGCAGCTGTGTTAAGATGGGAGAACTTCCAAAGACATTTGGTCATATTT 360
Db 672 CATCATGCAACAGTTAAGATGGTCAAGCTTCCAAAGACATTTGGTCATATTT 731
Qy 361 TTCACAGTACAGAAATCATGTGACATTTAATGACATCTTGAAAGTTCTGTAAAGA 420
Db 732 TTCACAGTACAGAAATCATGTGACATTTAATGACATCTTGAAAGTTCTGTAAAGA 791
Qy 421 TATAGTTCAATTAAGAACTTATGAAGAACTACAGAGTCAAGATTTGATGTTCT 480
Db 792 TGTAGTTCAATTAAGAACTTATGAAGAACTACAGAGTCAAGATTTGATGTTCT 851
Qy 481 TGCAGATGCTGTTTCCCTTTGGTGAAGCTGCTGGCGAGTTACTTAAATACCTTTGT 540
Db 852 TGCAGATGCTGTTTCCCTTTGGTGAAGCTGCTGGCGAGTTACTTAAATACCTTTGT 911
```

```
Qy 541 CTACAGCTCCGCTCTCTCTGAGCTACGCAATTTGAAAGCATAGTGAAGACTTCTGT 600
Db 912 GTACAGCTACGCTCTCTCTGAGCTACGCAATTTGAAAGCATAGTGAAGACTTCTGT 971
Qy 601 CCTCTCTCTATGAGCTGCTGTTATGTCAGAACTAAGTACCAATGACTTTCATAGA 660
Db 972 CCTCTCTCTATGAGCTGCTGTTATGTCAGAACTAAGTACCAATGACTTTCATAGA 1031
Qy 661 GAGGTTAAAAATATGATCTATGCTTATTTTGAATTTTGGTTCCAAATATTGACAT 720
Db 1032 GAGGTTAAAAATATGATCTATGCTTATTTTGAATTTTGGTTCCAAATATTGACAT 1091
Qy 721 GAAGAAGTGGATCAGTTCTTACAGTGAAGTTCTAGAA 758
Db 1092 GAAGAAGTGGATCAGTTCTTATGATGAAGTTCTAGAA 1129
```

```
RESULT 8
US-09-671-317-403
; Sequence 403, Application US/09671317
; Patent No. 6528260
; GENERAL INFORMATION:
; APPLICANT: Blumenfeld, Marla
; APPLICANT: Chumakov, Ilya
; APPLICANT: Bougueleret, Lydie
; APPLICANT: Cohen, Amick
; TITLE OF INVENTION: BIALLERIC MARKERS RELATED TO GENES INVOLVED IN DRUG METABOLISM
; FILE REFERENCE: 62.US.CIP
; CURRENT FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: US/09/671,317
; PRIOR FILING DATE: 2000-03-23
; PRIOR APPLICATION NUMBER: PCT/IB00/00403
; PRIOR FILING DATE: 2000-03-24
; PRIOR APPLICATION NUMBER: US 60/126,269
; PRIOR FILING DATE: 1999-03-25
; PRIOR APPLICATION NUMBER: US 60/131,961
; PRIOR FILING DATE: 1999-04-30
; NUMBER OF SEQ ID NOS: 977
; SOFTWARE: Patent.pm
; SEQ ID NO 403
; LENGTH: 1001
; TYPE: DNA
; ORGANISM: Homo Sapiens
; FEATURE:
; NAME/KEY: allele
; LOCATION: 503
; OTHER INFORMATION: 10-448-266 : polymorphic base A or C
; NAME/KEY: misc_binding
; LOCATION: 483..502
; OTHER INFORMATION: 10-448-266.misl, potential
; NAME/KEY: misc_binding
; LOCATION: 504..523
; OTHER INFORMATION: 10-448-266.mis2, potential complement
; NAME/KEY: primer_bind
; LOCATION: 238..257
; OTHER INFORMATION: upstream amplification primer
; NAME/KEY: primer_bind
; LOCATION: 660..679
; OTHER INFORMATION: downstream amplification primer, complement
; NAME/KEY: misc_binding
; LOCATION: 491..515
; OTHER INFORMATION: 10-448-266 potential probe
; NAME/KEY: misc_feature
; LOCATION: 772,907,915
; OTHER INFORMATION: n=a, g, c or t
US-09-671-317-403
```

```
Query Match 34.7%; Score 572.4; DB 4; Length 1001;
Best Local Similarity 86.4%; Pred. No. 1.7e-157;
Matches 643; Conservative 1; Mismatches 97; Indels 3; Gaps 1;
```

```

OY 15 AACGATTGCATTCGATGATGCTATGAAATGAGACTTCTCTCTGATACAGC 74
DB 2 AATTATCATTGACATGACAAAGATGCTCTGAAATGACTACAG---TTCTGTGATACAC 58
OY 75 TGAAGCTGTTACTTACCTGCTGGAAGTTGTGAAAAGTGTGCTGTGCTGCAAGAAATTCA 134
DB 59 TCAGATTTTACTTACCTGCTGGAAGTTGTGAAAAGTGTGCTGTGCTGCAAGAAATTCA 118
OY 135 GCCCACTGATGATATTAAGCAATCCCTGATGAACTTGTCCAGAGAGGTATAGAGGA 194
DB 119 GCCCTTGTGATGAAATGAAAGCAATCCCTGAAAGAACTTGTTCAGAGAGGTATAGAGGA 178
OY 195 CTGATTTGAGCACTTACAGCTTCCATTTCTTTCGATCCCAAGCCCACTCTCTTAAT 254
DB 179 CTGATTTGAGCACTTACAGCTTCCATTTCTTTCGATCCCAAGCCCACTCTCTTAAT 238
OY 255 TTGAAGTTTATCTGTATCTTTAACTAAACTGAGTTGAGATATTAACAAGACTGG 314
DB 239 TTGAAGTTTATCTGTATCTTTAACTAAACTGAGTTGAGATATTAACAAGACTGG 298
OY 315 TTAAGAGATGGGCAACTTCCAAAGACACATTTTGTCTATTTTTCAGAGTACAG 374
DB 299 TTAAGAGATTTGTCAAGAAATTCAAAAGATACATTTTGTCTATTTTTCAGAGTACAG 358
OY 375 AATCATGATGGACATTTATGACATCTTAGAAAGTTCTGTAAGATATAGTTCAATA 434
DB 359 AATCATGATGGACATTTATGACATCTTAGAAAGTTCTGTAAGATATAGTTCAATA 418
OY 435 AGAAACTTATGAGAACTACAGAGTCAAGATTTGATGTTGTTCTTTCAGAGTCTGTT 494
DB 419 AGAAACTTATGAGAACTACAGAGTCAAGATTTGATGTTGTTCTTTCAGAGTCTGTT 478
OY 495 TCCCTTTGATGAGTCTGCTGCGCCAGTACTTAAATCCCTTGTCTACAGCCCTCGCT 554
DB 479 TACCTGTGAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 538
OY 555 TCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 614
DB 539 TCACTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 598
OY 615 TCCCTGTTGTTATGTCAGAACTAGTGAACCAATGACTTTCATAGAGAGGTAAATA 674
DB 599 TACCTGTTGTTATGTCAGAACTAGTGAACCAATGACTTTCATAGAGAGGTAAATA 658
OY 675 TGATCTATGCTGCTTATTTGATTTTGTTCCAATATTTGACATGAAAGTGTGATC 734
DB 659 TGCTCTATGCTGCTTATTTGACTTTTGTTCCAATATTTGATGAAAGTGTGATC 718
OY 735 AGTCTACAGTGAAGTTCTAGAA 758
DB 719 AGTCTACAGTGAAGTTTCTAGAA 742

```

RESULT 9
US-09-356-806-114
Sequence 114, Application US/09356806
Patent No. 6586175

GENERAL INFORMATION:
APPLICANT: Penny, Laura
APPLICANT: Galvin, Margaret
APPLICANT: Miller, Andrew
APPLICANT: Reidy, Michael
TITLE OF INVENTION: Genotyping Human
TITLE OF INVENTION: UDP-Glucuronosyltransferase 2B4 (UGT2B4), 2B7 (UGT2B7) and
FILE REFERENCE: SEQ-22PRV2
CURRENT APPLICATION NUMBER: US/09/356,806
NUMBER OF SEQ ID NOS: 164
SOFTWARE: FaestSeq for Windows Version 3.0
SEQ ID NO 114
LENGTH: 2312
TYPE: DNA

```

ORGANISM: H. sapiens
FEATURE:
NAME/KEY: exon
LOCATION: (692) ... (1425)
US-09-356-806-114
Query Match 31.4%; Score 513; DB 4; Length 2312;
Best Local Similarity 80.6%; Pred. No. 6,2e-140;
Matches 613; Conservative 0; Mismatches 145; Indels 3; Gaps 1;
OY 1 AGCACTGGAAGAAACAGCACTTGCATTCAGATGCTATGAAATGAGCACTTCACTCT 60
DB 668 AACCACTGGAAGAAAGAACATTCATAGACAGAGATGCTGAAATGAGCACTGATCTT 727
OY 61 TCTCTGATACAGCTGATGCTGTTACTTACTCTGAGAGTTGGAAGGTGCTGTGTG 120
DB 728 TCTCTGATACAGCTGATGCTGTTACTTACTCTGAGAGCTGGAAGGTGCTGTGTG 787
OY 121 GCCCAAGAAATTCAGCCATGATGAATATTAAGACAAATCTGATGAACCTGTCCAG 180
DB 788 GCCCAAGAAATTCAGCCATGATGAATATTAAGACAAATCTGATGAAGCTGTGTCCAG 847
OY 181 AGTCTAGAGGTGATGCTATTTGCACTTCACTTCTTTCGATCCCAAGCC 240
DB 848 GGGTCATGAGGTGATGCTATTTGCACTTCACTTCTTTCGATCCCAAGCC 907
OY 241 ATCTACTCTTAAATTTGAAGTTTATCTGATCTTAACTTAACTGATGATGATAT 300
DB 908 ATCTACTTAAATTTGAAGTTTATCTGATCTTAACTTAACTGATGATGATAT 967
OY 301 TATCAAGCAGCTGTTAAGAGATGGCA--GAACTTCAAAGACACATTTGCTGATA 357
DB 968 TCTCTGAAATTTCTGATGATGATGATATATGATGATGATGATGATGATGATGATG 1027
OY 358 TTTTTCACAGTAAAGAAATCATGAGCACTTAAAGCACTTAAAGCTGTGAA 417
DB 1028 TTTTTCACAGTAAAGAAATCATGAGCACTTAAAGCACTTAAAGCTGTGAA 1087
OY 418 GGATATGTTTCAATATGAAGAACTTATGAAGAACTACAGAGTCAAGATTTGATGTGT 477
DB 1088 AGATGCACTTTGAAATGAAGAACTTATGATGAAGAACTACAGAGTCAAGATTTGATGT 1147
OY 478 TCTTGCAGATGCTGTTTCCCTTTGAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 537
DB 1148 TCTGCGAGATGCTGCTTATATCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1207
OY 538 TGTCTACAGCTCCGCTTCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 597
DB 1208 TGTCTACAGCTCCGCTTCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1267
OY 598 GTTCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 657
DB 1268 GTTCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1327
OY 658 AGAGAGGTAAGAAATATGATCTATGCTTATTTGATTTGATTTGCTCAAAATTTTGA 717
DB 1328 GGAAGGATTAAGAAATATGATCTATGCTTATTTGATTTGATTTGCTCAAAATTTTGA 1387
OY 718 CATGAAGAGTGGGATGCTGCTACAGTGAAGTTCTAGAA 758
DB 1388 TCTGAAGAGTGGGATGCTGCTACAGTGAAGTTCTAGAA 1428

```

RESULT 10
US-09-671-317-412
Sequence 412, Application US/09671317
Patent No. 6528260

GENERAL INFORMATION:
APPLICANT: Blumenfeld, Marta
APPLICANT: Chumakov, Ilya
APPLICANT: Bouguetel, Lydie
APPLICANT: Cohen, Amick
TITLE OF INVENTION: BIALLELIC MARKERS RELATED TO GENES INVOLVED IN DRUG METABOLISM

FILE REFERENCE: 62.US3.CIP
CURRENT APPLICATION NUMBER: US/09/671.317
CURRENT FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: US 09/536,178
PRIOR FILING DATE: 2000-03-23
PRIOR APPLICATION NUMBER: PCT/IB00/00403
PRIOR FILING DATE: 2000-03-24
PRIOR APPLICATION NUMBER: US 60/126,269
PRIOR FILING DATE: 1999-03-25
PRIOR APPLICATION NUMBER: US 60/131,961
PRIOR FILING DATE: 1999-04-30
NUMBER OF SEQ ID NOS: 977
SOFTWARE: Patent.pm
SEQ ID NO 412
LENGTH: 1001
TYPE: DNA
ORGANISM: Homo Sapiens
FEATURE:
NAME/KEY: allele
LOCATION: 503
OTHER INFORMATION: 10-457-284 : polymorphic base G or T
NAME/KEY: misc binding
LOCATION: 483..502
OTHER INFORMATION: 10-457-284.misl, potential
NAME/KEY: misc binding
LOCATION: 504..523
OTHER INFORMATION: 10-457-284.mis2, potential complement
NAME/KEY: primer bind
LOCATION: 220..238
OTHER INFORMATION: upstream amplification primer
NAME/KEY: primer bind
LOCATION: 621..639
OTHER INFORMATION: downstream amplification primer, complement
NAME/KEY: misc binding
LOCATION: 491..515
OTHER INFORMATION: 10-457-284 potential probe
NAME/KEY: misc feature
LOCATION: 715
OTHER INFORMATION: n=a, g, c or t
US-09-671-317-412

Query Match 30.4%; Score 502.2; DB 4; Length 1001;
Best Local Similarity 80.3%; Pred. No. 5.8e-137;
Matches 612; Conservative 1; Mismatches 145; Indels 4; Gaps 2;
QY 1 AGCACTGGAAACACACATTGTCATTCAGATGCTTATGAATGACCTGCT 60
DB 217 AACAACTGGAAAGAACATTTGCAVAAAGACAGATGCTCTGAAATGACGTCAGCTT 276
QY 61 TCTCTGATCAGCTGAGCTGTTACTTACGCTGGGAGTTGGGAAGGCTGGTG 120
DB 277 TCTGCTGATCAGCTGAGCTGTTACTTACGCTGGGAGTTGGGAAGGCTGGTG 336
QY 121 GCCCAGCAATTCAGCCATGATGAATATAAGACAACTCGGATGAATTCAGAG 180
DB 337 GCCCAGCAATTCAGCCATGATGAATATAAGACAACTCGGATGAATTCAGAG 336
QY 181 AGGTGATGAGTGAATGATGATGATGATGATGATGATGATGATGATGATGAT 240
DB 397 GGGGTGATGAGTGAATGATGATGATGATGATGATGATGATGATGATGATGAT 456
QY 241 ATCTACTTAAATTTGAATTTATCTGATCTTTAATTAAGATTTGAGATAT 300
DB 457 ATCTGCTAATTAATTAAGATTTATCTGATCTTTAATTAAGATTTGAGATAT 516
QY 301 TATCAAGCAGCTGGTTAAGATGGCA--GAATTCAAAAGACATTTGGTCTATA 357
DB 517 TCTTCTGAAATCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 576
QY 358 TTTTTCACAGTACAGAAATCATGTGACATTTAATGACATCTTGAAGTTCTGTAA 417
DB 577 TTTTTCACATTTACAGAAATTTGTTGGGAATTTATGACTACAGTAAACGCTGTAA 636

QY 418 GATATAGTTTCAATATAGAACTTATAGAGAACTACAGAGTCAAGATTTGATGTTGT 477
DB 637 AGATGACAGTTTGAATAGAACTTATAGAGAACTACAGAGTCAAGATTTGATGTTGT 696
QY 478 TCTTGCAGATGCTGTT-TTCCCTTTGGTGAAGTCTGCGCGGATTAATTAATACCT 536
DB 697 TCTGCGAGATGCTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 756
QY 537 TTGCTACAGCTCCGCTTCTCCGCTGCTACAGCAATTAAGATTAAGATTAAGATTA 596
DB 757 TTCTGACAGCTCTTCAATTTCTGTTGGCTACACATTTGGAAGATGGGAGATTC 816
QY 597 TGTTCCTCTCTCTATGCTGCTGTTTATGTCAGACTAAGTCAATGACCTTCA 656
DB 817 TGTTCCTCTCTCTATGCTGCTGTTTATGTCAGACTAAGTCAATGACCTTCA 876
QY 657 TGAAGAGGTAAATAATATGATCTATGCTTTATTTGATTTGTTCCAAATATTG 716
DB 877 TGAAGAGGTAAATAATATGATCTATGCTTTATTTGATTTGTTGTTCAATTTATG 936
QY 717 ACATGAAGATGGATGATCTTACAGTAACTTCTAGAA 758
DB 937 ATCTGAAGATGGGACAGTTTATATGAACTTCTAGGTA 978

RESULT 11

Sequence 2, Application PC/TUS9200282
GENERAL INFORMATION:
APPLICANT: OWENS, IDA S.
TITLE OF INVENTION: THE GENETIC LOCUS UGT1 AND A MUTATION
TITLE OF INVENTION: THEREIN.
NUMBER OF SEQUENCES: 40
CORRESPONDENCE ADDRESS:
ADDRESSEE: CUSHMAN DABRY & CUSHMAN
STREET: 1615 L STREET, N.W.
CITY: WASHINGTON
STATE: D.C.
COUNTRY: U.S.A.
ZIP: 20036-5601
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US92/00282
FILING DATE: 19920110
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: SCOTT, WATSON T.
REGISTRATION NUMBER: 26581
REFERENCE/DOCKET NUMBER: 91532-PCT
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-861-3000
TELEFAX: 202-822-0944
TELEX: 6714627 CUSH
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 2339 base pairs
TYPE: NUCLEIC ACID
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
PCT-US92-00282-2

Query Match 19.8%; Score 327.4; DB 5; Length 2339;
Best Local Similarity 52.9%; Pred. No. 1.2e-85;
Matches 806; Conservative 0; Mismatches 701; Indels 18; Gaps 4;
QY 65 CTGATCAGCTGAGCTGTTACTTACTGAGATTTGGAAGGCTGCTGGGCCC 124

46 CTGCTGCTCCTCCTCAGTGTCCAGCCCTGGGCTGAGTGAAGAGTGTGTGTCGCC 105
QY 125 ACAGATTCAGCCACTGATGAATATTAAGACAATCTGTGATGAACTTGTCCAGAGGT 184
DB 106 ACTGATGAGCCCTGCTCAGCATCGGAGGCTTGGGGGCTTCATGCCAGAGGC 165
QY 185 CATGAGGTGACTGATTGGGACTTTCAGCTTCCATTTCTTGCATCCCAAGCCCATCT 244
DB 166 CACGAGGCGGTGCTTCACCCAGAGTGAATGACATCAAGAGAGAAATTTTTC 225
QY 245 ACTCTTAATTTGAAGTTATCTGTATCTTTAACTAAAGATTGAGATATATC 304
DB 226 ACCCTGA-----CAGCTATGCTGTTCATGACCCAGAGAAATTTTCGCTTA-- 277
QY 305 AAGCAGCTGTTAAGATGAGGCGAAGCTTCCAAAGACATTTTGTCTATTTTCA 364
DB 278 ----CGCTGGGCTCAGTCAAGGGTTCCTTGAACAGAACTCTGGAAGATATCT 333
QY 365 CAAGTACAAGAAATCATGTGACATTTAATGACATCTTAAGAACTCTGTAAGATATA 424
DB 334 AGAAGTATGCAATTAATGAACAATGTATCTTGGCCCTTCATAGTGTGTGAGCTA 393
QY 425 GTTTCAAATAGAAGAACTTATGAGAACTACAGAGTCAAGATTTGATTTTGTGCA 484
DB 394 CTGATATATGAGGCGCTGATCAGACCTGATGCTACTTCTTGTATGTGTTTAA 453
QY 485 GATGCTGTTTCCCTTTGTGAGTGTGCTGCGAGTACTTAAATACCTTTGTCTAC 544
DB 454 GACCCCTTAACCTGCGGGGCGGTGCTGCTAATGCTGCTCAATTCGCTGTGTT 513
QY 545 AGCTCCGCTCTCTCCTGGGCTAGCAATTTGAAAAGCATGAGAGACTTGTTCCT 604
DB 514 TTTGAGAGTACATTCATGTGAC--TTAGCTTTAAGGACACAGTGTCCAAATCT 570
QY 605 CTTCTCTATGTCCTGTGTATGTACAGAACTAAGTGAACAATGACTTTGATAGAGG 664
DB 571 TCTCTCTATTTCTTAAGTACTAAGCAATTCAGACCATGACATTCCTGGAAGG 630
QY 665 GTAAAAATATGATCTATGTGCTTTATTTGAATTTTGTCCAAATTTTGAACATGAG 724
DB 631 GTCAAGAACATGCTCCTGCTGCGCTGCTCAATTTTGCACAACTTTTCTGCCCT 690
QY 725 AAGTGGATCAAGTCTACAGTGAAGTTCTAGAGAACCACTAAGTATCTGAGCAATG 784
DB 691 TATGCAAGTCT--TTCCTCTGAGCTTTTTCAGAGAGAGTGTCAAGTGTGATCTTGT 747
QY 785 GCAAAAGCTGACATATGCTTATTTGAAACTACTGGATTTTCAATTTCTTCCACCTC 844
DB 748 AGCTATGATCCGTGTGGCTGTTCGAGGGGACTTTGTGATGAGACTAACCCAGGCGATC 807
QY 845 TTACCAAAATGTTGAGTGTGTTGAGGACTCCACTGCAAACTGCAAACTTACCGAAG 904
DB 808 ATGCCCAACATGCTCTTCAATGAGGGGCAATCACTGTGCCAAGGGGAGCACTATCTCG 867
QY 905 GAAAGGAAGATTTGTCCAGAGCTCTGAGAGAAATGTTGTGTTTCTCTGCGGG 964
DB 868 GAATTTGAAGCTCAATTAATGCTCTTGAAGAACTGAGATTTGTTTCTTTTGGGA 927
QY 965 TCGATGTCATGTAACAGCTCAGAGAAAGGCGCAATGTAATGCAATCAGCCCTTGGCAAG 1024
DB 928 TCAATGTGTCTAGAAATTTCCAGAGAAAGGCTATGAGATGCTGATCTTTGGGCAAA 987
QY 1025 ATCCCAAAAGGTTCTGTGAGATTTGATGGGAATAAACAGATCTTTAGACTCAAT 1084
DB 988 ATCCCTCAGACAGTCTGTGGGCTGCTGAGAACCGACATGCAATCTTTCGAACAC 1047
QY 1085 ACTCGCTGTACAAGTGAATACCCAGAGTGAATCTTGTGATCCCAAAAGGCT 1144
DB 1048 AGGATATCTTTAAGGCTTACCCCAAAAGATGCTGTGTCACCCGATGACCGGTGC 1107
QY 1145 TTTATCACTCATGTGGAATGAATGAGATCTATGAAGCTATTTACATGGGTCTCTATG 1204
DB 1108 TTTATCACTCATGTGCTTCCATGTGTATTGAAGAACTATGCAATGGCGTTCCTCATG 1167

QY 1205 GTGGAGTTCCTCATATTTGGTGTATCAGCTTGATTAACATAGCTCATGAAGGCCAAAGGA 1264
DB 1168 GTATATGATGCCCTGTGTTGGTGTATCAGATGGAACAATGCAAAAGCGCATGAGACTAAGGA 1227
QY 1265 GCACTGTGAATAATTAACCTTCAAAACTATGACAGAGGAAGATTTACTGAGGCTTTGAGA 1324
DB 1228 GCTGAGATGACCTCGAATGTCTGGAATGACTTCTGAATATTTAAGAAATGCTTAA 1287
QY 1325 ACAGTATTCGATTCCTTATTAAGAGATGTATGATGATTAATCAAGATTCACAT 1384
DB 1288 GCACTCATATGACAAAGTTTCAAGAGAAACATCAGGCGCTCTCCAGCTTCAAG 1347
QY 1385 GATCAACTGTAAAGCCCTAGATGACAGAGCTTTGTGATGAGATTTGTATGCGCCAC 1444
DB 1348 GACCGCGGTGAGGCGCTGAGACTGAGCCGTGTTCTGGGTGAGTTGTATGAGGCGAC 1407
QY 1445 AAGAGCCAGACCTGCGATGAGTGCCTCACTGACCTGTTCCAGCACTACTCT 1504
DB 1408 AAGGCGCGCAGACCTGCGCCCGCAGCCAGCACTGATACAGTATCCATTC 1467
QY 1505 ATGATGTGATGGGCTCTGACCTGTGTGCAACTGTATATCTTGTTCACAAA 1564
DB 1468 TTGACGTGATGGTTCCTTGGCCGTGCTGTGACAGTGGCTTCAACCTTAA 1527
QY 1565 TGTTTTATTTTCTGTCAAAAT 1589
DB 1528 TGTGTCTATGTGCTACCGGAAT 1552

RESULT 12
PCT-US92-00282-1
Sequence 1, Application PC/TUS9200282
GENERAL INFORMATION:
APPLICANT: OWENS, IDA S.
APPLICANT: RITTER, JOSEPH K.
TITLE OF INVENTION: THE GENETIC LOCUS UGT1 AND A MUTATION
TITLE OF INVENTION: THEREIN.
NUMBER OF SEQUENCES: 40
CORRESPONDENCE ADDRESS:
ADDRESSER: CUSHMAN DARBY & CUSHMAN
STREET: 1615 L STREET, N.W.
CITY: WASHINGTON
STATE: D.C.
COUNTRY: U.S.A.
ZIP: 20036-5601
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US92/00282
FILING DATE: 19920110
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: SCOTT, WATSON T.
REGISTRATION NUMBER: 26581
REFERENCE/DOCKET NUMBER: 91532-PCT
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-861-3000
TELEFAX: 202-822-0944
TELEX: 6714627 CUSH
INFORMATION FOR SEQ ID NO. 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2336 base pairs
TYPE: NUCLEIC ACID
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
PCT-US92-00282-1
Query Match 19.2%; Score 316; DB 5; Length 2336;

Best Local Similarity 55.2%; Pred. No. 2.5e-82;
Matches 660; Conservative 0; Mismatches 530; Indels 6; Gaps 2;

```

QY 394 TGACATCTAGAAAGTTCTGTAGATATAGTTTCAAAATAGAACTTAAAGAACT 453
Db 360 TGCTATCTTTTGTCTGCTGTTCCCACTTACACAAAGAGCTCATGCGCTCCCT 419
QY 454 ACAGAGTCAAGATTTGATGTTCTTGTGAGATGCTGTTTCCCTTGGTGAGTCT 513
Db 420 GGCAGAAAGAGCTTTATGTCATGCTGAGGACCTTCTCTTCTTGACCCCATCT 479
QY 514 GCGCGATTAATAAAATACCCTTTGTCTACAGCTCCGCTTCTCTCGCTACGCAAT 573
Db 480 GCGCCATACCTGCTCTGCGCACTGATTTCTTTCATGACATGCGCATGACAC--CT 536
QY 574 TGAAGACATAGTGGAGAGCTCTGTTCCCTCTTCTATGCTGCTGTATGTCAGA 633
Db 537 GGAATTTGAGGCTTACCCAGTCCCAACCAATTCCTTACGCTCCAGGCTCTCTCTC 596
QY 634 ACTAAGTACCAATGATCTTCTATAGAGAGGTTAAATAATGATCTATGCTTATTT 693
Db 597 TCATTCGATCATGATGCTTCTCTGACGGGTGAAGACATGCTCATTTGCTTTTACA 656
QY 694 TGAATTTGTTCCAAATTAATTGACATGAAGAGTGGATCAATCTACAGTGAATCT 753
Db 657 GAACTTCTGTGCGACGTGTTTATCCCGTATGCAACCTTGCTCA---GAATTCCT 713
QY 754 AGGAAGCCCATCTCTATCTGAGACATGCGAAAGTGCATGATGCTTATTCGAAA 813
Db 714 TCAGAGAGAGGTGCTGTCCAGACCTATGAGCTGTGCTGTGCGGTGTTTAAAG 773
QY 814 CTACTGGATTTTCAATTTCTCTCACCCACTCTTACCAATGTTAGTTGTTGAGAGCT 873
Db 774 TGACTTTGTAAAGATTAACCTTAGGCCATCATCCCAATATGTTTGTGTTGGAAT 833
QY 874 CCACTGCAAACTGCGCAAACTCCCTACCGAAAGAAATGAAAGTTGTCCAGAGCTCTG 933
Db 834 CAATGCTCTTACCAAAATCACTATCCAGAAATTTGAAGCTTCACTTATGCTTCTGG 893
QY 934 AGAAAAATGTTGTTGTTGTTTCTCTGGGTGATGTCATGTAACGTCAGAAAG 993
Db 894 AGAAACATGAAATGTGTTTCTTCTTGGATCAATGATCTCAGAAATTCAGAGAA 953
QY 994 GGCATATGTAATTCATCAGCCCTTGCAGAGATCCCAAAAGTTCTGTGAGATTGA 1053
Db 954 AGCTATGCAATTCCTATGCTTTGGGCAAAATCCTTACACATCCTGTCGCGTACAC 1013
QY 1054 TGGGAATMAACAGATATCTTAGATCTCAATCTCGGCTGTACAAAGTGAATCCCGAAA 1113
Db 1014 TGGAAACCGAACATCGAATCTTGGGAAACACAGATCTGTTTAAAGTGGCTACCCAAA 1073
QY 1114 TGATCTTTGTTGATCTCCCAAAACCAAGCTTTTATCATCTCANTGTGAAATGAA 1173
Db 1074 CGATCTGCTTGTGACCCGATGACCCGCTTATATACCATGCTGTTCCCATGCTGT 1133
QY 1174 CTATGAAGTATTTTACATGAGGCTCTATGTTGGAGTTCCCATTTTGTGATGAGCT 1233
Db 1134 TTATGAAGCATATGCAATGAGCGTTCCCATGATGATGATGCTTGTGTTGATGAT 1193
QY 1234 TGAATCATGCTCACTGAAGGCGCAAGAGAGAGCTGTGAAATAAATCTCAAACTAT 1293
Db 1194 GGAATATGCAAGAGGATGAGGAGTAAAGGAGCTGAGTACCTGATGTTCTGGAAT 1253
QY 1294 GACAAGGAAAGATTTACTGAGGCTTTGAGAACATGATTAACGATTTCTCTTATAA 1353
Db 1254 GACTTCTGAAGATTTTAAATAATGCTTAAAGAGCATCATGACAAAAGTTTACAAGA 1313
QY 1354 GAATGCTATGATTAACAAGATTTCAAGATGATCAACCTGTAAAGCCCTAGATGAGC 1413
Db 1314 GAACATAGGCGCTCTTCAGCTTCAAGAGACCGCCGTTGAGCGCTGAGCTGGC 1373
QY 1414 AGTCTTGTGATGAGTTTGTGATGCGCAAAAGAGCAAGAGCACTGCGATGAGCTGC 1473

```

```

Db 1374 CGTGTCTGAGTGGATTTGTGTATGAGGCAAGAGGCGCCACACTGCGCCGAGC 1433
QY 1474 CCATGACCTCAGCTGGTTCCAGACTACTATAGATGATGATGGTTCTGCTGACCTG 1533
Db 1434 CCAGACCTCAGCTGGTACAGATGACATGATCTTGGAGATGATGATGCTTCTGGCGCT 1493
QY 1534 TGTGCACTGCTATATCTTGTTCACAAATGTTTATTTTCTGTGCAAAAT 1589
Db 1494 CGTGTGACAGTGGCTTATCATCTTAAATGTTGTGCTTATGAGCTACCGGAAT 1549

```

RESULT 13

```

US-09-305-856B-17
; Sequence 17, Application US/09305856B
; Patent No. 6479236
; GENERAL INFORMATION:
; APPLICANT: Penny, Laura
; APPLICANT: Galvin, Margaret
; TITLE OF INVENTION: Genotyping the Human
; TITLE OF INVENTION: UDP-Glucuronosyltransferase 1 (UGT1) Gene
; FILE REFERENCE: 4389-7 (formerly SEO-17C1P)
; CURRENT APPLICATION NUMBER: US/09/305,856B
; CURRENT FILING DATE: 1999-05-05
; PRIOR APPLICATION NUMBER: 60/084,807
; PRIOR FILING DATE: 1998-05-07
; NUMBER OF SEQ ID NOS: 124
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 17
; LENGTH: 735
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(735)
US-09-305-856B-17

```

Query Match 16.3%; Score 269; DB 4; Length 735;

Best Local Similarity 62.0%; Pred. No. 7.9e-69;

Matches 425; Conservative 0; Mismatches 260; Indels 0; Gaps 0;

```

QY 905 GAAATGAAAGATTTGTCCAGAGCTCTGAGAAATGTTGTTGTTTCTCTGGGG 964
Db 1 GAAATGAAAGATTTGTCCAGAGCTCTGAGAAATGTTGTTGTTTCTCTGGGG 60
QY 965 TCGATGTCATGTAACAGTCAAGAAAGAGGCGCAATGTAATGTCATGAGCTTGGC 1024
Db 61 TCAATGTCATGTAACAGTCAAGAAAGAGGCGCAATGTAATGTCATGAGCTTGGC 120
QY 1025 ATCCCAAAAGGTTCTGTGAGATTTGATGGAATTAACCAATGATCTTAAAGACTCAAT 1084
Db 121 ATCCCTCAGACAGTCTGTGAGGCTGACCTGACATGGAACCGACATGATCTTGGACAC 180
QY 1085 ACTCGGCTGTAAGTGAATACCCAGAAATGATCTTCTGATGATCCCAAAACCAAGCT 1144
Db 181 ACATATCTGTTAAAGTGGCTACCCCAAAAGATCTGTTGATGATGATGATGATGATG 240
QY 1145 TTTATCACTATGTTGATGATGATGATGATGATGATGATGATGATGATGATGATG 1204
Db 241 TTTATCACTATGTTGATGATGATGATGATGATGATGATGATGATGATGATGATG 300
QY 1205 GTGAGAGTTCCATATTTGTTGATGATGATGATGATGATGATGATGATGATGATGATG 1264
Db 301 GTGATATGCTCTTGTGTTGTTGATGATGATGATGATGATGATGATGATGATGATGATG 360
QY 1265 GAGCTGTGAAATTAATCTTCAAACTATGATGACAAAGGAAATTTACTGAGGCTTTGAGA 1324
Db 361 GCTGAGATGACCTGATGTTCTGAAATGATCTTGAAGATTTGAAGATTTGCTTAA 420
QY 1325 ACAGTATTAACGATCTCTTTTAAAGAGATGATGATGATGATGATGATGATGATGATG 1384
Db 421 GAGTATCATTAATACAAAGTTTAAAGAGAAATGATGATGATGATGATGATGATGATG 480
QY 1385 GATCAACCTGTAAAGCCCTGATGAGAGCTTCTGATGATGATGATGATGATGATGATG 1444

```

| | | | |
|----|------|--|------|
| Db | 481 | GACCGCCCGGAGACCGCTGACCTTGACCCTGTTCTTGAGATTGTGATGAGCAC | 540 |
| Oy | 1445 | AAAGAGCCAAGCACTCGATACAGCTGCCATGACCTCACCTGGTCCAGACACTACT | 1504 |
| Db | 541 | AAGGCGCGCCACACCTTGCGCCCGGACGCCACAGACTCACTGTATACAGTACCATTC | 600 |
| Oy | 1505 | ATAGATGATTTGGGTTCTCTGTGACCTGTGTGGCACTGCTATATCTTGTTCAAAA | 1564 |
| Db | 601 | TTGACGCGATTTGGTTTCTCTCTTGCCCGCTCGTGTGACAGTGGCCTTATCACCTTAA | 660 |
| Oy | 1565 | TGTTTTTATTTTCTCTGTCAAAAT | 1589 |
| Db | 661 | TGTTGCTTATGCTACCGAAAT | 685 |

RESULT 14
 US-09-677-317-427
 ; Sequence 427, Application US/09671317
 ; Patent No. 6528260
 ; GENERAL INFORMATION:
 ; APPLICANT: Blumenfeld, Marta
 ; APPLICANT: Chumakov, Ilya
 ; APPLICANT: Bougueleret, Lydie
 ; APPLICANT: Cohen, Annick
 ; TITLE OF INVENTION: BIALLELIC MARKERS RELATED TO GENES INVOLVED IN DRUG METABOLISM

```

Query Match      14.8%; Score 245; DB 4; Length 1001;
Best Local Similarity 85.8%; Pred. No. 9,66-62;
Matches 272; Conservative 0; Mismatches 45; Indels 0; Gaps 0

442 TATGAAGAAACATCAGAGAGTCAAGATTGATGTGTTCTTGACAGATCGTTTCCCTT 501
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
QY

```

| | | | |
|----|-----|--|-----|
| Db | 1 | TATGAGAAACTACAAAGAGTCAAAATTTGATGTCCCTTCGGGAGATGCGGTTAAATCCCG | 50 |
| Qy | 502 | TGGTGAAGCTGCTGGCGCGAGTACTTAAATATACCCCTTGTCTACAGCCTCGGCTCTCTCC | 561 |
| Db | 61 | TGGTGAAGCTGCTGGCTGAGCTACTTAAACATACCTCTTCTGTACAGCTCCGCTCTCTGT | 120 |
| Qy | 562 | TGGCTACGCAATTGAAAAGCATAGTGAAGACTTGTTCCTCCTCTCTCTAATGTGCTGT | 621 |
| Db | 121 | TGGCTACACAGTGTGAGAAATGTGTGAAGATTTCTGTCCCTCTCTCTAATGTACTGT | 180 |
| Qy | 622 | TGTTATGTCAGATCAATGAGCCAAATGACTTTCATAGAGAGGTAAAAAATATGATCTA | 681 |
| Db | 181 | TGTTATGTCAGATTTAAGTATCAATGATTTTCAATGAGAGATTAATAAATATGATATA | 240 |
| Qy | 682 | TGTGCTTAATTTGATATTTGGTTCGAAATATTTGACATGAAGAAGTGGGATCACTTCTA | 741 |
| Db | 241 | TATGCTTAATTTTGTACTTTTGGTTTCAAGCATATATCTGAAGAAAGTGGGACCAAGTTTTA | 300 |
| Qy | 742 | CAGTGAAGTCTAGGAA | 758 |
| Db | 301 | TAGTGAAGTCTTAGGTA | 317 |

RESULT 15
US-09-671-317-428
: Sequence 428, Application US/09671317

```

GENERAL INFORMATION:
APPLICANT: Blumentfeld, Marta
APPLICANT: Chumakov, Ilya
APPLICANT: Bougueleret, Lydie
APPLICANT: Cohen, Annick
TITLE OF INVENTION: BIALLISTIC MARKERS RELATED TO GENES INVOLVED IN DRUG METABOLISM
FILE REFERENCE: 62.US3.CIP
CURRENT APPLICATION NUMBER: US/09/671,317
CURRENT FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: US 09/536,178
PRIOR FILING DATE: 2000-03-23
PRIOR APPLICATION NUMBER: PCT/IB00/00403
PRIOR FILING DATE: 2000-03-24
PRIOR APPLICATION NUMBER: US 60/126,269
PRIOR FILING DATE: 1999-03-25
PRIOR APPLICATION NUMBER: US 60/131,961
PRIOR FILING DATE: 1999-04-30
NUMBER OF SEQ ID NOS: 977
SOFTWARE: Patent.pm
SEQ ID NO 428
LENGTH: 1001
TYPE: DNA
ORGANISM: Homo Sapiens
FEATURE:
NAME/KEY: allele
LOCATION: 503
OTHER INFORMATION: 12-621-49 : polymorphic base A or G
NAME/KEY: misc_binding
LOCATION: 483..502
OTHER INFORMATION: 12-621-49.misl, potential
NAME/KEY: misc_binding
LOCATION: 504..523
OTHER INFORMATION: 12-621-49.mis2, potential complement
NAME/KEY: primer_bind
LOCATION: 455..473
OTHER INFORMATION: upstream amplification primer
NAME/KEY: primer_bind
LOCATION: 907..927
OTHER INFORMATION: downstream amplification primer, complement
NAME/KEY: misc_binding
LOCATION: 491..515
OTHER INFORMATION: 12-621-49 potential probe
NAME/KEY: misc_feature
LOCATION: 634
OTHER INFORMATION: n=a, g, c or t
US-09-671-317-428

```


| | | | | |
|-----------------------|--------------|--------------------|----------------|--------------|
| Query Match | 14.7%; | Score 242.8; | DB 4; | Length 1001; |
| Best Local Similarity | 76.8%; | Pred. No. 4.2e-61; | | |
| Matches 322; | Conservative | 0; | Mismatches 93; | Indels 4; |
| | | | | Gaps 2; |

QY 1 AGCAATCGGAAAACAGCAATGTGCATTGCATCAGATGTCATATGAAATGGAATCT-TCAGTCC 59

Db 583 AGCGATTTGGAAAAGAACATTGCATTGCACCGAGAACTATGAAATGATGATGCAATA 642ATGTCGAATA

QY 60 TTCTCTCGATACAGCTGAGCTGTACTTATTAGCTCTGGAGATTTGGAAAAGGTGCTGTGT 119

Db 643 TTCTCTCGATACAACTCAGTTGTCATTATGTTCTGGAGTTGTGGAAAAGGTGCTAGTGT 702

QY 120 GGGCCACAGAAATTCAGCCACTGGATGATATATTAAGACAATCTTGATTAATTTGTCCAGA 1797

Db 703 GGGCCACAGAAATCACCCCAATTTGATGTAATATTAAGACAATCTTGATTAATTTGTTCGA 762

QY 180 GAGGTCAATGAGAGTGTATTTGGCATCTTCAGTCTTCATTTCTTGATCCCAACAGCC 239

Db 763 GAGGCCATTAAGATGACTGTCTGTCAATCTTTGGCTTCATTTATTTACCTCTGTAAAT 822

QY 240 CATCTACTCTTAAATTTGAAGTTATCTGTATCTTTAACTTAACAATGAGTTGAGGATA 299

Db 823 CATCTGCTATTAGATTTGAAGTTATCCAACTCTAATTAATAAATTTTAAAGGTC 882

QY 300 TTATCAAGCAGCTGTGTTAAGATGGGCA--GAACTTCCAAAAGACATTTTGGTCAT 356

Db 883 TTGTGTGAAACGATCAATAGATGATTAAGATCTTCAGAAAGATCATTTTGGTCAT 942

QY 357 ATTTTTCACAAGTACAGAAATCATGTGACATTTAAGACATCTTGAAAGTTCTGT 415

Db 943 ATTTCACACAGACAGAAATCTTTTGGGAATCTACAGACTGTGTTAATATATCTCTGT 1001

Search completed: December 7, 2003, 07:39:50
Job time : 95 secs

THIS PAGE BLANK (USPTO)


```

Db 124 GCCACAGAAATTCAGCCACTGGATGAATATAAGACAATCTGGATGAATCTTGCAGAG 183
Qy 181 AGGCATGAGGTGATCTGATTTGGCATCTTCAGCTTCCATTTCTTCCATCCCAACAGCC 240
Db 184 AGGCATGAGGTGATCTGATTTGGCATCTTCAGCTTCCATTTCTTCCATCCCAACAGCC 243
Qy 241 ATCTACTCTTAAATTTGAAGTTTATCTGATCTTAACTTAACTGATTTGAGATAT 300
Db 244 ATCTACTCTTAAATTTGAAGTTTATCTGATCTTAACTTAACTGATTTGAGATAT 303
Qy 301 TATCAAGCAGCTGTTAAGAGATGGGCAACCTTCCAAAGACACATTTTGGTCATAT 360
Db 304 TATCAAGCAGCTGTTAAGAGATGGGCAACCTTCCAAAGACACATTTTGGTCATAT 363
Qy 361 TTCAAGATACAAAGAAATCATGTGACATTTAATGACATCTTGAAGAAAGTTCTGAAGA 420
Db 364 TTCAAGATACAAAGAAATCATGTGACATTTAATGACATCTTGAAGAAAGTTCTGAAGA 423
Qy 421 TATAGTTTCAATTAAGAACTTATGAAGAACTACAGAGTCAAGATTTGATTTGTTCT 480
Db 424 TATAGTTTCAATTAAGAACTTATGAAGAACTACAGAGTCAAGATTTGATTTGTTCT 483
Qy 481 TGCAGATGCTGTTTCCCTTTGGTGAAGCTGCGCGAGTTACTTAAATACCTTTGT 540
Db 484 TGCAGATGCTGTTTCCCTTTGGTGAAGCTGCGCGAGTTACTTAAATACCTTTGT 543
Qy 541 CTACAGCCTCCGCTTCTCTCTGCTGACGCAATTTGAAGACATAGTGAAGACATTTCT 600
Db 544 CTACAGCCTCCGCTTCTCTCTGCTGACGCAATTTGAAGACATAGTGAAGACATTTCT 603
Qy 601 CCCCTCTCTCTATGAGCTGTTGTTATGTGCAAGCTAAGTGAACCAATGATTTCTATGA 660
Db 604 CCCCTCTCTCTATGAGCTGTTGTTATGTGCAAGCTAAGTGAACCAATGATTTCTATGA 663
Qy 661 GAGGTAAGAAATATGATCTATGCTTATTTGAATTTGGTTCCAAATATTTGACAT 720
Db 664 GAGGTAAGAAATATGATCTATGCTTATTTGAATTTGGTTCCAAATATTTGACAT 723
Qy 721 GAAAGATGGGATCACTTCTACAGTGAAGTTCTAGAAAGCCCACTACGTTATCTGAGAC 780
Db 724 GAAAGATGGGATCACTTCTACAGTGAAGTTCTAGAAAGCCCACTACGTTATCTGAGAC 783
Qy 781 AATGGCAAAAGCTGACATATGAGCTTATGCAAGCTACGAGATTTTCAATTTCTCCACC 840
Db 784 AATGGCAAAAGCTGACATATGAGCTTATGCAAGCTACGAGATTTTCAATTTCTCCACC 843
Qy 841 ACTCTTACCAAAATGTTGAGTTGTTGAGAGCTCCACTGCAAACTGCAAACTCCCTAAC 900
Db 844 ACTCTTACCAAAATGTTGAGTTGTTGAGAGCTCCACTGCAAACTGCAAACTCCCTAAC 903
Qy 901 GAAAGAAATGGAAGATTTGTCAGAGCTCTGAGAGAAATGTTGTTGTTCTCT 960
Db 904 GAAAGAAATGGAAGATTTGTCAGAGCTCTGAGAGAAATGTTGTTGTTCTCT 963
Qy 961 GGGGTGATGTCAGTAAACAGTCAAGAAAGGCAATTAATGATGATAGCCCTTGC 1020
Db 964 GGGGTGATGTCAGTAAACAGTCAAGAAAGGCAATTAATGATGATAGCCCTTGC 1023
Qy 1021 CAAGATCCACAAAGGTTCTGTGAGATTTGATGGAATTAACAGATCTTTAGACT 1080
Db 1024 CAAGATCCACAAAGGTTCTGTGAGATTTGATGGAATTAACAGATCTTTAGACT 1083
Qy 1081 CAATACTGCTGTACAAAGTGAATCCCAAGATGATCTTCTGTCATCCCAAAACCA 1140
Db 1084 CAATACTGCTGTACAAAGTGAATCCCAAGATGATCTTCTGTCATCCCAAAACCA 1143
Qy 1141 AGCTTTATCACTAGTGTGAATGAATGAGATCTATGAAGCTATTTACATGGGGTCCC 1200
Db 1144 AGCTTTATCACTAGTGTGAATGAATGAGATCTATGAAGCTATTTACATGGGGTCCC 1203
Qy 1201 TATGATGAGGATCCCATATTTGATGATGATGATGATGATGATGATGATGATGATGAT 1260

```

```

Db 1204 TATGATGAGGATCCCATTTGTTGAGATCAACCTGATMACTTGACACATGAAGCCAA 1263
Qy 1261 AGGAGCAGCTGTGAAATTAACCTTCAAACTATGACAGAGGATTTACGAGGCTTT 1320
Db 1264 GGGAGCAGCTGTGATTTGAGATTTCCACAAATGTGAGTACAGACTTACATGCACT 1323
Qy 1321 GAGAACAGTATTAACCGATCTCTTATTAAGAAATGCTATGATGATTAATCAAGATTTCA 1380
Db 1324 GAGAACAGTATTAATGATCTCTTATTAATTAAGAAATGCTATGATTAATCAAGATTTCA 1383
Qy 1381 CATGATCAACTGTAAAGCCCTAGATGAGCAGCTTCTGATGAGTGTGTCATGCG 1440
Db 1384 TCATGATCAACTGTAAAGCCCTAGATGAGCAGCTTCTGATGAGTGTGTCATGCG 1443
Qy 1441 CCACAAAGGAGCAAGACCTGAGATGAGTGCCTCATGACCTCACTGTTCCAGACTA 1500
Db 1444 CCATAAAGGAGCAAGACCTGAGGTTGAGGCCCAAGACCTCACTGTTCCAGACTA 1503
Qy 1501 CTCTATGATGATGATTTGGTTCTGCTGACCTGTTGAGCAACTGCTATATTTCTTTTAC 1560
Db 1504 CTCTTGGATGATGATGATTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1563
Qy 1561 AAAATGTTTATTTTCTGCTGCAAAAATTTAATAAATTAAGAAAGTGAAGAAAGAGGA 1620
Db 1564 AAAATGCTGTTTGTGCTG---GAAGTTGTTAGAAACAGAAAGGAGGAAAGAGA 1620
Qy 1621 ATGATCTTTC 1631
Db 1621 TTAATTAGCTC 1631

RESULT 3
US-10-057-834A-1
; Sequence 1, Application US/10057834A
; Publication No. US2003009960A1
; GENERAL INFORMATION:
; APPLICANT: RATNIN, MARK J.
; APPLICANT: INNOCENTI, FEDERICO
; APPLICANT: DAS, SOMA
; APPLICANT: IYER, LALITHA
; APPLICANT: SAWYER, MICHAEL
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR OPTIMIZING UGT2B7 SUBSTRATE DOSINGS
; FILE REFERENCE: ARCD-358US
; CURRENT APPLICATION NUMBER: US/10/057, 834A
; PRIOR FILING DATE: 2002-08-22
; PRIOR APPLICATION NUMBER: UNKNOWN
; PRIOR FILING DATE: 2002-01-25
; NUMBER OF SEQ ID NOS: 78
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 1991
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (151)..(1740)
; US-10-057-834A-1

Query Match 77.1%; Score 1272.6; DB 14; Length 1991;
Best Local Similarity 86.3%; Pred. No. 0;
Matches 1407; Conservative 0; Mismatches 224; Indels 0; Gaps 0;

```

237 GGACAGCAGAAATACAGCCATTGGATGATATTAAGCAATCTGTGATGACTTATTCAGAG 296
QY 181 AGGTGATAGAGTGAATGATGAGCATCTTCAGCTCCATTTCTTGATGCCAAGGCC 240
Db 297 AGGTATAGAGGATGATGAGCATCTTCAGCTCCATTTCTTGATGCCAAGGCC 356
QY 241 ATCTACTCTTAAATTTGAAGTTTATCTGTATCTTAACTAAAGCTGATTTGAGAGATAT 300
Db 357 ATCCGCTCTTAAATTTGAAGTTTATCCCAATCTTAACTAAAGCTGATTTGAGAGATAT 416
QY 301 TATCAGAGAGCTGTTAAGAGATGGGCAAGACTTCCAAAGACACATTTGGTCAATTT 360
Db 417 CATCATGCAACAGATTTAAGAGATGGTCAAGCTTCCAAAGATATCTTTGGTCAATTT 476
QY 351 TTCAAGATGACAGAAATCATGTGATGCAATTTAATGACATCTTAAAGTTCTGTAAGA 420
Db 477 TTCAAGATGACAGAAATCATGTGATGCAATTTGATGACATCTTAAAGTTCTGTAAGA 536
QY 421 TATAGTTTCAATTAAGAAATCTTATGAGAACTACAGAGTCAAGATTTGATGTTCT 480
Db 537 TGTATTTCAATTAAGAAATTTAAGAAAGTACAGAGTCAAGATTTGATGTTCT 596
QY 481 TGCAGATGCTGTTTCCCTTGGTGGAGCTGCTGCGCAATTTAAATACCTTTGT 540
Db 597 TGCAGATGCTATTTTCCCTGAGTGGAGCTGCTGCGCAATTTAAATACCTTTGT 656
QY 541 CTACAGCTCGGCTCTCTCCGCTGAGCAATTAAGAACTAGTGGAGGCTCTGTT 600
Db 657 GTACAGCTCGAGCTCTCTCCGCTGAGCAATTTAAGAACTAGTGGAGGCTCTGTT 716
QY 601 CCTCTCTCTCTATGCTGCTGTTGATGATGACAACTAAGTACCAATGACTTTGATAGA 660
Db 717 CCTCTCTCTCTATGCTGCTGTTGATGATGACAACTAAGTACCAATGACTTTGATAGA 776
QY 661 GAGGTAATAAATATGATCTATGCTTATTTGAATTTGGTCCAAATTTGATCAT 720
Db 777 GAGGTAATAAATATGATCTATGCTTATTTGAATTTGGTCCAAATTTGATCAT 836
QY 721 GAAGAAGTGGATGATCTACAGTGAAGTCTAGGAAGACCCACTAGTATCTGAGAC 780
Db 837 GAAGAAGTGGATGATCTACAGTGAAGTCTAGGAAGACCCACTAGTATCTGAGAC 896
QY 781 AATGCAAAAGCTGACATATGCTTATTTGAAATCTAGTGGATTTCAATTTCTCACCC 840
Db 897 AATGCAAAAGCTGACATATGCTTATTTGAAATCTAGTGGATTTCAATTTCTCACCC 956
QY 841 ACTCTTACCAATGTTGAGTTGCTGTTGAGAGACTCCACTGCAAACTGCCCAACCCCTAC 900
Db 957 ACTCTTACCAATGTTGAGTTGCTGTTGAGAGACTCCACTGCAAACTGCCCAACCCCTAC 1016
QY 901 GAAGAAGTGGATGATCTACAGTGAAGTCTAGGAAGTGGTGGTGGTGGTGGTGGTGGT 960
Db 1017 TAAAGAAATGGAAGCTTTGTACAGAGCTCTGGAAGAAATGATGTTGGTGGTGGTGGTGGT 1076
QY 961 GGGGTGATGCTCAGTAAACAGCTGAGAAAGGCAATGTAATTTGATCAGCCCTTGC 1020
Db 1077 GGGGTGATGCTCAGTAAACAGCTGAGAAAGGCAATGTAATTTGATCAGCCCTTGC 1136
QY 1021 CAAGATCCCAAAAGTTCTGTGAGATTTGATGGAATTAACAGATACTTTAGACT 1080
Db 1137 CCAGATCCCAAAAGTTCTGTGAGATTTGATGGAATTAACAGATACTTTAGACT 1196
QY 1081 CAATCTGCGCTGTCAAGTGAATCCCAAGATGATCTTCTGTCATCCCAAAACCA 1140
Db 1197 CAATCTGCGCTGTCAAGTGAATCCCAAGATGATCTTCTGTCATCCCAAAACCA 1256
QY 1141 AGCTTTTATCACTATGCTGAGATGAATGAGATCTATGAGATATTTACATGAGGCTCC 1200
Db 1257 AGCTTTTATCACTATGCTGAGATGAATGAGATCTATGAGATATTTACATGAGGCTCC 1316
QY 1201 TATGTTGGAGATTTCCCATATTTTGTGATCAGTTGATTAACATAGTCAATGAGGCCAA 1260
Db 1317 TATGTTGGAGATTTCCCATATTTTGTGATCAGTTGATTAACATAGTCAATGAGGCCAA 1376

QY 1261 AGAGCAGCTGTAGAAATTAATCAAACTATGACAGGAGGATTTAAGAGGCTTT 1320
Db 1377 GGGAGCAGCTGTAGAGTGAATGATCTTCAACAAATTCAGATACAGTCTGCTGATTCAT 1436
QY 1321 GAGAACATCATTAACCATTCCTCTTATTAAGAGATGCTATGAGATTTATCAAGATTTCA 1380
Db 1437 GAAGAGATTAATTAATGATCCTTCATATTAAGAGATGCTATGAGATTTATCAAGATTTCA 1496
QY 1381 CCATGATCAACCTGTAAAGCCCTGATGAGAGCTCTGATGAGTGGATTTGTCAGCG 1440
Db 1497 ACATGATCAACCATGTAAGCCCTGATGAGAGCTCTGATGAGTGGATTTGTCAGCG 1556
QY 1441 CCAGAAAGAGCCAGAGCAGCTGAGATGAGTCCATGATGAGTCCATGCTGTTCCAGACATA 1500
Db 1557 CCAGAAAGAGCTTAACACCTTGGGTTGAGAGCCAGACCTGATCCTGTTCCAGATNCA 1616
QY 1501 CTCTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1560
Db 1617 CTCTTGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1676
QY 1561 AAAATGTTTATTTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1620
Db 1677 AAAATGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1736
QY 1621 ATAGATCTTTC 1631
Db 1737 TTAGTATATTC 1747

RESULT 4
US-09-981-353-193
; Sequence 193, Application us/09981353
; Patent No. US20020160382A1
; GENERAL INFORMATION:
; APPLICANT: Lasek, Amy W.
; TITLE OF INVENTION: GENES EXPRESSED IN COLON CANCER
; FILE REFERENCE: PA-0038 US
; CURRENT APPLICATION NUMBER: US/09/981,353
; NUMBER OF SEQ ID NOS: 194
; SOFTWARE: PERL Program
; SEQ ID NO 193
; LENGTH: 1714
; TYPE: DNA
; ORGANISM: Homo sapiens
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. US20020160382A1 088078CB1
US-09-981-353-193

Query Match 77.0%; Score 1271.2; DB 10; Length 1714;
Best Local Similarity 86.3%; Pred. No. 0;
Matches 1405; Conservative 0; Mismatches 223; Indels 0; Gaps 0;

QY 4 AACTGGAAGAAACAGCATTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 63
Db 1 AACTGGAAGAAACAGCATTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 60
QY 64 CTTGATCAGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 123
Db 61 GCTATATCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 120
QY 124 CACGAATTCAGCACTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 183
Db 121 AGCAGAAATACAGCATTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 180
QY 184 TCAGAGGAT 243
Db 181 TCAGAGGAT 240
QY 244 TACTCTTAATTTGAAGTTTATCTGATATCTTAACTAAAGTGAAGTGGATATATAT 303

| | | | |
|----|------|---|------|
| Db | 241 | CGCTTTAAATTTGAAATTTATCCACACTCTTTAACTAAACTGAGTTGGAGAAATTCAT | 300 |
| Qy | 304 | CAAGCAGCTGGTTAAGAGATGGGCAGAACTTCCAAAAGACAATTTGGTCATATTTTTC | 363 |
| Db | 301 | CATCGAACAGATTTAAGAGATGGTCAGACCTTCCAAAAGATCACTTTGGTTATATTTTC | 360 |
| Qy | 364 | ACAAGTACAAGAATCANTGGACATTTTAATGACATCTTGAAGAAAGTCTGTAAAGAT | 423 |
| Db | 361 | ACAAGTACAGAAATCATGTCAATTTTGGTACATTACTGAAGAAAGTCTGTAAAGAT | 420 |
| Qy | 424 | AGTTTCAATTAAGAACTTATGAAGAACTACAGAGATCAAGATTTGATGTGTTCTGC | 483 |
| Db | 421 | AGTTTCAATTAAGAAATTTATGAAAAAGTACAGAGATCAAGATTTGACGTTCATTTTTC | 480 |
| Qy | 484 | AGATGCTGTTTTCCCTTTGGTGGAGCTGTGGCCAGTTACTTAAATACCTTTGTCTA | 543 |
| Db | 481 | AGATGCTATTTTTCCCTGTAGTGAAGCTGTGGCTAGCTATTTAACATACCTTTGTGTA | 540 |
| Qy | 544 | CAGCTCCGCTTCTCTCTGGCTACGCAATTTGAAAAAGATAGTGGAGACCTTCTGTCC | 603 |
| Db | 541 | CAGTCTCAGCTTCTCTCTGGCTACCTTTTGAAGAAAGATAGTGGAGATTTATTTTCC | 600 |
| Qy | 604 | TCCTTCCATGNGCCTGTTGTTATGTCAAGACTAAGTACCAAAATGACTTTTCATAGAG | 663 |
| Db | 601 | TCCTTCTTACGTACTGTTGTTATGTCAAGAAATTACTATCAAAATGACTTTTCATAGAG | 660 |
| Qy | 664 | GGTAAAAAATATGATCTATGTGCTTTATTTTGAATTTTGGTCCAAATATTTTGAATGAA | 723 |
| Db | 661 | GGTAAAAAATATGATCTATGTGCTTTACTTTGACTTTTGGTTCGAAATATTTTGAATGAA | 720 |
| Qy | 724 | GAAGTGGGATCAATTTCTACAGTGAAGTTCTAGAGAAAGCCCACTACGTTATCTGAGCAAT | 783 |
| Db | 721 | GAAGTGGGATCAATTTTATAGTGAAGTTCTAGAGAAAGCCCACTACGTTATCTGAGCAAT | 780 |
| Qy | 784 | GGCAAAAGCTGACATATAGGCTTTATTCGAAACATACTGGGATTTTCAATTTCTCCACCACT | 843 |
| Db | 781 | GGGAAAGCTGACATATAGGCTTTATTCGAAACCTCTGGAAATTTTCAATTTCTCCACCACT | 840 |
| Qy | 844 | CTTACCAAAATGTTGAGTTCGTTGAGAGACTCCACTGCAAACTGCGCAAACTCTTACCGAA | 903 |
| Db | 841 | CTTACCAAAATGTTGATTTTGTTTGAGAGACTCCACTGCAAACTGCGCAAACTCTTACCGTAA | 900 |
| Qy | 904 | GGAATGGAAGATTTGTCCAGAGCTCTGAGAGAAAAATGTTGTTGTGTTTCTCTGGG | 963 |
| Db | 901 | GGAATGGAAGACTTTGTACAGAGCTCTGAGAGAAAAATGTTGTTGTGTTTCTCTGGG | 960 |
| Qy | 964 | GTGATGGTCACTAACACGTACGAAGAAAGGGCCAAATTAATTTGATACAGCCCTTGGCAA | 1023 |
| Db | 961 | GTCAATGGTCACTAACATGACGAAGAAAGGGCCAACTAAATTTGATACAGCCCTTGGCCCA | 1020 |
| Qy | 1024 | GATCCCAAAAGGTTCTGTGGAGATTGATGGGAATTAACACAGATCTTTTAGAGACTCAA | 1083 |
| Db | 1021 | GATCCCAAAAGGTTCTGTGGAGATTGATGGGAATTAACACAGATCTTTTAGGTTCTCAA | 1080 |
| Qy | 1084 | TACTGGGCTGTACAAGTGGATACCCACGAATGATCTTGTGTCAATCCCAAAACCAAGC | 1143 |
| Db | 1081 | TACTGGGCTGTACAAGTGGATACCCCAAAATGACCTTCTAGGTCAATCCAAAGACAGAGC | 1140 |
| Qy | 1144 | TTTTTATCCTATGTTGGTGAATGAATGGGATCTATGAAAGCTATTTTACATGGGGTCCAT | 1203 |
| Db | 1141 | TTTTTATACCTATGTTGGTGAAGCAATGGGATCTATGAAAGCAATTCATATGGGATCCCAT | 1200 |
| Qy | 1204 | GGTGGGAGTTCCCAATTTGGTGAATGAGTGTAACTATGCTCACATGAAGGCCAAAGG | 1263 |
| Db | 1201 | GGTGGGAGTTCCATTTGTTGCTGATCAACCTGATACCTTGTCTCAATGAAGGCCAAGGG | 1260 |
| Qy | 1264 | AGCAGCTGTAGAAATTAACCTTCAAAATATGACAAAGCGAAGATTTTACTGAGGGCTTTGAG | 1323 |
| Db | 1261 | AGCAGCTGTATAGAGGAGACTTCAACACATGTGAGTACAGAGCTTGTGATGATTTCAAGATTTCA | 1320 |
| Qy | 1324 | AACAGTATTACCAATTTCTCTTATTAAGAGATGCTATGAGATTTATCAAGATTTCA | 1383 |

| | | | |
|---|------|---|------|
| Db | 1321 | GAGGTAATTAAATGATCTCTTCATATATAAGAAATGTTATGAATTTATCAAGAAATTCACA | 1389 |
| Qy | 1384 | TGATCAACCTGTAAGCCCTAGATGCAGCAGTCTTTGGATGCAGTTTGTTCATGCGCA | 1443 |
| Db | 1381 | TGATCAACCAAGTGAAGCCCTCGATTCAGCAGTCTTTGGATGAATTTGTTCATGCGCA | 1440 |
| Qy | 1444 | CAAGGAGCAAGCAGCTGGATGAGCTGAGCTGACCAGTCACTCAGTGTTCAGCAGCTCTC | 1503 |
| Db | 1441 | CAAGGAGCTTAACACCTTGCGGTTGACGACCCAGACCTCACTCGTGTTCAGTACACTC | 1500 |
| Qy | 1504 | TATAGATGATGTTGGTTCCTGCTGACCTGTGTGCGCAACTGCTATATCTTGTTCACAA | 1563 |
| Db | 1501 | TTTGGATGATGTTGGTTCCTGCTGCTGTGTGTGCAACTGTGATATTTATTCGTACAAA | 1560 |
| Qy | 1564 | ATGTTTTTATATTTTCTGCTCAAAAATTTAATTAAGAAAGATAGAAAAAGGGAATA | 1623 |
| Db | 1561 | ATGTTGCTGTGTTTGTCTTTCGGAAGTTTGCTAGAAAAAGAAAGGAAAAATGATTA | 1620 |
| Qy | 1624 | GATCTTC 1631 | |
| Db | 1621 | GTTATATC 1628 | |
| RESULT 5 | | | |
| US-09-880-107-2120 | | | |
| ; Sequence 2120, Application US/09880107 | | | |
| ; Patent No. US20020142981A1 | | | |
| ; GENERAL INFORMATION: | | | |
| ; APPLICANT: Horne, Darci T. | | | |
| ; APPLICANT: Vockley, Joseph G. | | | |
| ; APPLICANT: Scherf, Uwe | | | |
| ; APPLICANT: Gene Logic, Inc. | | | |
| ; TITLE OF INVENTION: Gene Expression Profiles in Liver Cancer | | | |
| ; FILE REFERENCE: 44921-5028-WO | | | |
| ; CURRENT APPLICATION NUMBER: US/09/880,107 | | | |
| ; CURRENT FILING DATE: 2001-06-14 | | | |
| ; PRIOR APPLICATION NUMBER: US 60/211,379 | | | |
| ; PRIOR FILING DATE: 2000-06-14 | | | |
| ; PRIOR APPLICATION NUMBER: US 60/237,054 | | | |
| ; PRIOR FILING DATE: 2000-10-02 | | | |
| ; NUMBER OF SEQ ID NOS: 3950 | | | |
| ; SOFTWARE: PatentIn Ver. 2.1 | | | |
| ; SEQ ID NO 2120 | | | |
| ; LENGTH: 1855 | | | |
| ; TYPE: DNA | | | |
| ; ORGANISM: Homo sapiens | | | |
| ; FEATURE: | | | |
| ; OTHER INFORMATION: Genbank Accession No. US20020142981A1 J05428 | | | |
| US-09-880-107-2120 | | | |
| Query Match | | | |
| Best Local Similarity 86.1%; Pred. No. 0; | | | |
| Matches 1387; Conservative 0; Mismatches 224; Indels 0; Gaps 0; | | | |
| Qy | 21 | TGCATTGCATGAGATGTCATGAATGAACTTCAGCTCTTCTCTGATACAGCTGAGCT | 80 |
| Db | 1 | TGCATTGCACACAGATGTCGTGAATGAACTTCAGTAAATTTGCTAATCAACTGAGCT | 60 |
| Qy | 81 | GTTACTTAACTCTGGGAGTTTGGAAAGGCTGCTGTGGCCCAAGAAATTCAGCACT | 140 |
| Db | 61 | TTTGCTTTAACTCTGGGAAATTTGGAAAGGCTGCTGTGGCCCAAGAAATTCAGCACT | 120 |
| Qy | 141 | GGATGAATATAAGACAATCTGTGATGAATCTTGTCCAGAGAGTCAATGAGTGACTGAT | 200 |
| Db | 121 | GGATGAATATAAGACAATCTGTGATGAGCTTATTCAGAGAGTCAATGAGTGACTGAC | 180 |
| Qy | 201 | TGGCATTCAGCTTCATTTCTTTGATCCCAACAGCCCAATCTACTCTTAATTTGAAG | 260 |
| Db | 181 | TGGCATTCAGCTTCATTTCTTTGATCCCAACATCATCGCTCTTAATTTGAAA | 240 |
| Qy | 261 | TTTATTCCTGATCTTTTAACATAAAGCTGATGAGATATTTATCAAGCAGCTGTTAAGA | 320 |
| Db | 241 | TTTATTCCTGATCTTTTAACATAAAGCTGATGAGATTTATCATATGCAACAGATTAAGA | 300 |

QY 321 GATGGGAGACTCCAAAGACATTTTGTGCTATTTTTCACAGTACAGAAATCA 380
DB 301 GATGGTACACCTTCCAAAGATATTTTGTGCTATTTTTCACAGTACAGAAATCA 360
QY 381 TGTGACATTTTAATGACATCTAGTAAAGTCTGTAAAGATATAGTTTCAATAGAAAC 440
DB 361 TGTCAATTTTGTGACATCTAGTAAAGTCTGTAAAGATATAGTTTCAATAGAAAT 420
QY 441 TTAAGAGAACTACAGAGTCAAGATTTGTGTTCTTTCAGATGCTGTTTCCCT 500
DB 421 TTAAGAGAAAGTCAAGATTTGTGATTTTTCAGATGCTGTTTCCCT 480
QY 501 TTGGGACCTGCTGGCCGAGTTACTTAATACCTTGTCTACAGCTCCGCTCTCTC 560
DB 481 GTAGGACCTGCTGGCTAGCTATTTACATACCTTGTGTGACGCTTCTCTCTC 540
QY 561 CTGGCTACGCAATTTGAAAGCATAGTGGAGACTTCTCTCTCTCTCTATGTGCTG 620
DB 541 CTGGCTACGCAATTTGAAAGCATAGTGGAGACTTCTCTCTCTCTCTATGTGCTG 600
QY 621 TTGTATGTCAAGTAAAGTCAAGTACTTTCATAGAGAGGTAAATATGATCT 680
DB 601 TTGTATGTCAAGTAAAGTCAAGTACTTTCATAGAGAGGTAAATATGATCT 660
QY 681 ATGCTCTTATTTTGAATTTTGTGCTCAAAATTTTGAAGTAAAGTGGAGTCT 740
DB 661 ATGCTCTTATCTTACTTGTGCTCAAAATTTTGAAGTAAAGTGGAGTCT 720
QY 741 ACAGTGAATTTTGAAGTAAAGTCAAGTACTTTCATAGAGAGGTAAATATGATCT 800
DB 721 ATAGTGAATTTTGAAGTAAAGTCAAGTACTTTCATAGAGAGGTAAATATGATCT 780
QY 801 GAGCTTATTTGAAGTAAAGTCAAGTACTTTCATAGAGAGGTAAATATGATCT 860
DB 781 GAGCTTATTTGAAGTAAAGTCAAGTACTTTCATAGAGAGGTAAATATGATCT 840
QY 861 TCGTGGAGAGTCCAGTGGAGTAAAGTCAAGTACTTTCATAGAGAGGTAAATATGATCT 920
DB 841 TGTGTTGAGAGTCCAGTGGAGTAAAGTCAAGTACTTTCATAGAGAGGTAAATATGATCT 900
QY 921 TCCAGAGTCTGGAGTAAAGTCAAGTACTTTCATAGAGAGGTAAATATGATCT 980
DB 901 TACAGAGTCTGGAGTAAAGTCAAGTACTTTCATAGAGAGGTAAATATGATCT 960
QY 981 CGTCAAGAGTAAAGTCAAGTACTTTCATAGAGAGGTAAATATGATCT 1040
DB 961 TGACAGAGTAAAGTCAAGTACTTTCATAGAGAGGTAAATATGATCT 1020
QY 1041 TGTGAGATTTTGAAGTAAAGTCAAGTACTTTCATAGAGAGGTAAATATGATCT 1100
DB 1021 TGTGAGATTTTGAAGTAAAGTCAAGTACTTTCATAGAGAGGTAAATATGATCT 1080
QY 1101 GGATACCCAGAGTAAAGTCAAGTACTTTCATAGAGAGGTAAATATGATCT 1160
DB 1081 GGATACCCAGAGTAAAGTCAAGTACTTTCATAGAGAGGTAAATATGATCT 1140
QY 1161 GAATGAATTTTGAAGTAAAGTCAAGTACTTTCATAGAGAGGTAAATATGATCT 1220
DB 1141 GAGCCAAATTTTGAAGTAAAGTCAAGTACTTTCATAGAGAGGTAAATATGATCT 1200
QY 1221 TTGGATTTTGAAGTAAAGTCAAGTACTTTCATAGAGAGGTAAATATGATCT 1280
DB 1201 TTGGATTTTGAAGTAAAGTCAAGTACTTTCATAGAGAGGTAAATATGATCT 1260
QY 1281 ACTTCAAACTTGAAGTAAAGTCAAGTACTTTCATAGAGAGGTAAATATGATCT 1340
DB 1261 ACTTCAAACTTGAAGTAAAGTCAAGTACTTTCATAGAGAGGTAAATATGATCT 1320
QY 1341 CCTTATTAAGAGTAAAGTCAAGTACTTTCATAGAGAGGTAAATATGATCT 1400
DB 1321 CCTTATTAAGAGTAAAGTCAAGTACTTTCATAGAGAGGTAAATATGATCT 1380

QY 1401 CCTTATTAAGAGTAAAGTCAAGTACTTTCATAGAGAGGTAAATATGATCT 1460
DB 1381 CCTTATTAAGAGTAAAGTCAAGTACTTTCATAGAGAGGTAAATATGATCT 1440
QY 1461 TGGGATGAGTCCAGTAAAGTCAAGTACTTTCATAGAGAGGTAAATATGATCT 1520
DB 1441 TGGGATGAGTCCAGTAAAGTCAAGTACTTTCATAGAGAGGTAAATATGATCT 1500
QY 1521 TCCGCTGAGCTGGGCAAGTCAAGTACTTTCATAGAGAGGTAAATATGATCT 1580
DB 1501 TCCGCTGAGCTGGGCAAGTCAAGTACTTTCATAGAGAGGTAAATATGATCT 1560
QY 1581 GTCAAAATTTTGAAGTAAAGTCAAGTACTTTCATAGAGAGGTAAATATGATCT 1631
DB 1561 TGTGAAATTTTGAAGTAAAGTCAAGTACTTTCATAGAGAGGTAAATATGATCT 1611

RESULT 6
US-10-205-522-39
; Sequence 39, Application US/10205522
; Publication No. US2003077629A1
; GENERAL INFORMATION:
; APPLICANT: Penny, Laura
; APPLICANT: Galvin, Margaret
; APPLICANT: Miller, Andrew
; APPLICANT: Reidy, Michael
; TITLE OF INVENTION: Genotyping Human
; TITLE OF INVENTION: UDP-Glucuronosyltransferase 2B4 (UGT2B4), 2B7 (UGT2B7) and
; TITLE OF INVENTION: 2B15 (UGT2B15) Genes
; FILE REFERENCE: SEQ-22PRV2
; CURRENT APPLICATION NUMBER: US/10/205,522
; CURRENT FILING DATE: 2002-07-24
; PRIOR APPLICATION NUMBER: US/09/356,806
; PRIOR FILING DATE: 1999-07-20
; NUMBER OF SEQ ID NOS: 164
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 39
; LENGTH: 1854
; TYPE: DNA
; ORGANISM: H. sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (15) ... (1584)
US-10-205-522-39

Query Match 75.6%; Score 1247.8; DB 14; Length 1854;
Best Local Similarity 85.9%; Pred. No. 0;
Matches 1384; Conservative 0; Mismatches 227; Indels 0; Gaps 0;

QY 21 TGCATTCATGAGATGCTATGAAATGAGTCTTCTCTGATACAGTGAAGT 80
DB 1 TGCATTCATGAGATGCTATGAAATGAGTCTTCTCTGATACAGTGAAGT 60
QY 81 GTTACTTACTGCTGGAGATGCTGAAAGTGTGCTGTGCTGCTGCTGCTGCTGCT 140
DB 61 TTTGCTTACTGCTGGAGATGCTGAAAGTGTGCTGTGCTGCTGCTGCTGCTGCT 120
QY 141 GATGATATTAAGCAATCTGATGAATCTTGTCCAGAGAGTCAATGAGTGAAT 200
DB 121 GATGATATTAAGCAATCTGATGAATCTTGTCCAGAGAGTCAATGAGTGAAT 180
QY 201 TGGCATCTTCAAGCTTCCATTTTTCATGCTCCCAAGCCCATCTTCTTAAATTTGAAG 260
DB 181 TGGCATCTTCAAGCTTCCATTTTTCATGCTCCCAAGCCCATCTTCTTAAATTTGAAG 240
QY 261 TTTATCTGATATCTTAACTGAAATGAGATTTATCAAGAGCTGTGAAGA 320
DB 241 TTTATCTGATATCTTAACTGAAATGAGATTTATCAAGAGCTGTGAAGA 300
QY 321 GATGGGCAAGCTTCCAAAGACATTTTGTGCTATTTTTCACAAATTAAGATCA 380
DB 301 GATGGTCAAGCTTCCAAAGACATTTTGTGCTATTTTTCACAAATTAAGATCA 360

| | | | |
|----|------|--|------|
| Qy | 381 | GTGGACATTTAATGACATCTACTAGAAAGTCTGTAAGAAATATAGTTTCAATAATAAGAAC | 440 |
| Db | 361 | TGTCAATATTTGGTGCATTAACCTAATAAAGTTCTGTAAGAAGTATGTTTCAATAATAAGAAAT | 420 |
| Qy | 441 | TTATGAANAACCTACAGAGTCAGAGTTTGAATGTTGTTCTTGACAGATGCTGTTTTCGCCCT | 500 |
| Db | 421 | TTATGAANAAGTACAGAGTCAAGATTTGAAGTCAATTTTGACAGATGCTATTTTTCCT | 480 |
| Qy | 501 | TTGGTAGCTGCTGGCCGAGTTTACTTAAATATCCCTTGTGTACAGCTCCGCTTCTCTC | 560 |
| Db | 481 | GTAGTAGCTCTGGCTGAGCTATTTAATATCCCTTGTGTACAGCTCCGCTTCTCTC | 540 |
| Qy | 561 | CTGGTACGCAATTGAAAAGCATAGTAGAGCTCTGTCCCTCCCTCTATGTGCTG | 620 |
| Db | 541 | CTGGCTACCTTTTGAAAAGCATAGTAGAGCTTATATTTTCCCTCTCTACGTAAGCTG | 600 |
| Qy | 621 | TTGTTATGTCAAGACTAAGTACCAATATGACTTTTCATAGAGAGGTAAATAATATGATCT | 680 |
| Db | 601 | TTGTTATGTCAAGATTAATGATTCAAATGACTTTTCATAGAGAGGTAAATAATATGATCT | 660 |
| Qy | 681 | ATGTGCTTATTTTGAATTTTGGTTCCAATATTTTGACATGAAGAAGTGGATCAGTTCT | 740 |
| Db | 661 | ATGTGCTTATCTTTGACTTTTGGTTCCAATATTTTGAATGATGAAGAAGTGGATCAGTTT | 720 |
| Qy | 741 | ACAGTGAAGTCTTAGAAGACCCCACTAGCTATCTGAACAATGGCAAAAGCTGCATAT | 800 |
| Db | 721 | ATAGTGAAGTCTTAGAAGACCCCACTAATATCTGAACAATGGGAAAGTGCATAT | 780 |
| Qy | 801 | GGCTTATTCGAAACTACTGGGATTTTCAATTTCCACCCACTCTTACCAATGTGTAGT | 860 |
| Db | 781 | GGCTTATTCGAACTCTTGAAATTTTCAGTTTCCATATCCACTCTTACCAATGTGTAGT | 840 |
| Qy | 861 | TCGTTTGAGAGCTCCACTGCAAACTGCGCAAAACCCCTACCGAAGAAATGGAAGATTTG | 920 |
| Db | 841 | TTGTTTGAGAGCTCCACTGCAAACTGCGCAAAACCCCTGCTTAAGAAATGGAAGCTTTG | 900 |
| Qy | 921 | TCAGAGCTCTGAGAAAATGTGTGTTGTGTGTTTCTCTGGGGTGCATGTGCTAACA | 980 |
| Db | 901 | TACAGAGCTCTGAGAAAATGTGTGTTGTGTGTTTCTCTGGGGTCAATGTGCTAACA | 960 |
| Qy | 981 | CGTCAAGAAAGAGGCCCAATGTAATTTGATCAGCCCTTGCCCAAGATGCCCAAAAAGTTT | 1040 |
| Db | 961 | TGACAGAAAGAAAGGCCCAAGTAAATTGATCAGCCCTTGCCCAAGATGCCCAAAAAGTTT | 1020 |
| Qy | 1041 | TGTGAGATTTGATGGAATAAACCCAGATCTTTAGACTAATCTCGGCTGTACAAGT | 1100 |
| Db | 1021 | TGTGAGATTTGATGGAATAAACCCAGATCTTTAGACTAATCTCGGCTGTACAAGT | 1080 |
| Qy | 1101 | GGATATCCCGAATATGATCTTCTTGCTATCCCAAAAACCAAGCTTTATCTCACTAGGTT | 1160 |
| Db | 1081 | GGATATCCCGAATAATGACCTTCTAGGTCATCCAAAAGACAGAGCTTTTATCTCATGTGTT | 1140 |
| Qy | 1161 | GAATGAATGGGATCTATAGCTATTTTACATGGGTCCTATGTTGGGAATGCCAAT | 1220 |
| Db | 1141 | GAGCCAAATGGGATCTACAGAGCAATCTACATGGGATCCCTATGTTGGGAATGCCAAT | 1200 |
| Qy | 1221 | TTGGTGAATCAGCTTGATTAACATAGTCTACATGAAGGCCAAAGAGACAGCTGTAGAAATTA | 1280 |
| Db | 1201 | TTGGCGATCAACCTGTATTAACATGTCTACATGAAGGCCAAGGAGACAGCTGTAGAGATG | 1260 |
| Qy | 1281 | ACTTCAAAAATATGACAAAGCAAGATTTTACTGAGGGCTTTGAGAAACAGTCAATTCAGAT | 1340 |
| Db | 1261 | ACTTCAACACATGTGCGATGACAGACTTGCTGAATGCAATGAAGAGATTAATTAATGATC | 1320 |
| Qy | 1341 | CCTCTTAATAAGAAATGCTATAGATTTATCAAGAAATTCACATGATCAACTGTAAAGC | 1400 |
| Db | 1321 | CTTCAATATAAAGAGATGTATGAATTAATCAAGAAATTCACATGATCAACAGTGAAGC | 1380 |
| Qy | 1401 | CCCTTAGATCGAGAGCTCTTCTGGAATCGAGTTTGTATCTCGCCACAAAGAGCCAAAGCAC | 1460 |
| Db | 1381 | CCCTTGATCGAGAGCTCTTCTGGAATTTGTTGATATGCTCACCACAAAGAGACTTAACACC | 1440 |
| Qy | 1461 | TGCGATCAGCTGCCCATGACCTCACCTGGTTCACAGACTACTTATAGATGTGATGGGT | 1520 |

| | | | |
|----|------|--|------|
| Db | 1441 | TTGGGGTTGCAGGCCCAAGACTCACTCGTGTTCCAGTACCACCTCTTTGGATGTATTTGGGT | 1500 |
| Qy | 1521 | TCCTGCTGACCTGTGTGGCAACTGCTATATTTCTTGTCACAAAATGTTTTTAATTTTCT | 1580 |
| Db | 1501 | TCCTGCTGCTGTGTGTGGCAACTGTGATATTATTCGACAAAATGTTGCTGTTTGGT | 1560 |
| Qy | 1581 | GTCAAAAATTTAATAAAGCTAGAAAGATAGAAAAAGAGGAATGATCTTTTC | 1631 |
| Db | 1561 | TCCTGGAAGTTTGCTAGAAAAGCAAGAAAGGAAAAATGATTACTTTATATC | 1611 |

RESULT 7

```

US-10-198-846-13134
: Sequence 13134, Application US/10198846
: Publication No. US2003009974A1
: GENERAL INFORMATION:
: APPLICANT: Little, James
: APPLICANT: Xu, Yongyao
: APPLICANT: Wang, Youzhen
: APPLICANT: Steilmann, Kathleen
: TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS
: TITLE OF INVENTION: FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND
: TITLE OF INVENTION: THERAPY OF BREAST CANCER
: FILE REFERENCE: MRI-049
: CURRENT APPLICATION NUMBER: US/10/198,846
: CURRENT FILING DATE: 2002-07-18
: PRIOR APPLICATION NUMBER: 60/306,220
: PRIOR FILING DATE: 2001-07-18
: NUMBER OF SEQ ID NOS: 14084
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 13134
: LENGTH: 2844
: TYPE: DNA
: ORGANISM: Homo sapiens
: FEATURE:
: NAME/KEY: misc feature
: LOCATION: 2834, 2835, 2836, 2837, 2828, 2829, 2830, 2831, 2832, 2833
: LOCATION: 2834, 2835, 2836, 2837, 2838, 2839, 2840, 2841, 2842, 2843
: LOCATION: 2844
: OTHER INFORMATION: n = A,T,C or G
US-10-198-846-13134

```

| | | | | |
|--------------|--------|---------------|--------|--------------|
| Query Match: | 75.0%; | Score 1237.2; | DB 14; | Length 2844; |
|--------------|--------|---------------|--------|--------------|

Best Local Similarity 85.5%; Pred.No. 0;
Matches 137; Conservative 0; Mismatches 233; Indels 0; Gaps 0;

Qy 22 GCATTCACACAGATGCTCATGAAATGCACTTCCTCTGATACAGCTGACCTG 81
Db 21 GCATTCACACAGATGACTTGAAATGCACTTCAGCTTCCTGCTATACATCTCACTG 80
Qy 82 TTACCTTAACTCTGGAGAGTTGTGGAAAGGTCTGGTGTGCCCCACAGAAATTCAGCCACTG 141
Db 81 TTACTTTTACCTCTGGGAGAGTTGTGGAAAGAGTCTGGTGTGGGCCCCACAAATACAGCCACTTG 140
Qy 142 GATGATATATAAAGACAATCTCGAGATGAACTGTCCAGAGAGGTGATAGAGGACTGATTT 201
Db 141 GATGATATATGAAAGACAATCTGAAAGACTGTTCACAGAGGTGATAGAGGTGACTGACT 200
Qy 202 GGCATCTTCAGCTCCATCTTTCTTTGATCCCAACAGCCCATCTACTCTTAATTTGAAGT 261
Db 201 GGCATCTTCAGCTCCATCTTTTGTGATCCCAATGATGCATCCACTCTTAATTTGAAGT 260
Qy 262 TTATCTCTGATCTTTTAACTTAAACCTGAGTTGAGAGATTTTACAAGCAGCTGTAAAG 321
Db 261 TTATCTCTCATCTTTTAACTTAAACCTGAAATTTGAAATATCATCAGCAAGGTTAAAG 320
Qy 322 ATGGCAGAACTTCCAAAAGACACATTTTGTGCATATTTTTCACAAGTACAGAAATCAT 381
Db 321 ATGCTGAGCAATTCGAAAAGATGCTTTTGTGTTATTTTTTCACAAAGACAGAAATCCT 380
Qy 382 GTGCACTTTAATGACATACCTTAGAAAGTCTGTAGGATATAGCTTTCAATTAAGAACT 441

381 GTGGAAATTATATGACATATTAGAACTCTGTAAAGATGATTCATAAATAGAAAGT 440
442 TATGAGAACTACAGAGTCAAGATTGATGTTGTTCTTGACAGATGCTGTTCCCTT 501
441 TATGAAAAAATACAGAGTCAAGATTTGACATGTTTTGAGATGCTGTTTTCCCTG 500
502 TGTGAGCTGCTGCGCGAGTTACTTAAATAATCCCTTGTCTACAGCTCCGCTTCTCC 561
501 TGTGAGCTGCTGCGCGCTACTTAACTACACCGTTTGTGACAGTCTCCGCTTCTCC 560
562 TGGCTACCAATTTGAAAAAGCATGTGAGAGACTTGTCTCCCTCTCTATGTCCTGT 621
561 TGGCTACCAATTTGAAAAAGCATGTGAGAGACTTGTCTCCCTCTCTATGTCCTGT 620
622 TGTATGTCAGAACCTAAGTACCAATGATCTTCTTACAGAGGGTAAAAAATATGATCTA 681
621 TGTATGTCAGAACCTAAGTACCAATGATCTTCTTACAGAGGGTAAAAAATATGATCTA 680
682 TGTGCTTATTTTGAATTTTGGTTCCTCAATAATTTGACATGAAAGTGGGATCAGTTCTA 741
681 TGTGCTTATTTTGAATTTTGGTTCCTCAATAATTTGACATGAAAGTGGGATCAGTTCTA 740
742 CAGTGAAGTCTAGAGAAAGCCCACTAAGTATCTGAGACAAATGSCAAAAGCTGACATATG 801
741 CAGTGAAGTCTAGAGAAAGCCCACTAAGTATTTGAGACAAATGSCAAAAGCTGACATATG 800
802 GCTTATTTGAAAATCTAGGATTTTCAATTTCTTCCACCACTCTAATCAATGTTGAGT 861
801 GCTTATTTGAAAATCTAGGATTTTCAATTTCTTCCATCATCTTCAACAAAGTGTGATTT 860
862 CGTTGAGGAGCTCAGCTCAAACTGCAAACTCCCTACCGAAGAAATGAAAGATTTGT 921
861 TGTGAGAGATTTCCACTGCAAACTGCAAACTCCCTACCGAAGAAATGAAAGATTTGT 920
922 CCAGAGCTCTGAGAAAAATGTTGTGTTGTTTCTCTGAGGCTGATGTCAGTACAC 981
921 ACAGAGCTCTGAGAAAAATGTTGTGTTGTTTCTCTGAGGCTGATGATGATACAT 980
982 GTCAAGAAAGGGCCAAATGTAATTTGATCAGCCCTTGCAAGATCCCAAAAGGTTCT 1041
981 GACACAGAAAGGGCCAAATGTAATTTGATCAGCCCTTGCAAGATCCCAAAAGGTTCT 1040
1042 GTGAGATTTGAGGAAATTAACAGAAATCTTGAAGTCAATACCTGCTGTACAAATG 1101
1041 GTGAGATTTGAGGAAATTAACAGAAATCTTGAAGTCAATACCTGCTGTACAAATG 1100
1102 GATACCCCAAGATGATCTTCTGTTGTCATCCAAAACAAAGCTTTTATCACTCATGTGG 1161
1101 GATACCCCAAGATGATCTTCTGTTGTCATCCAAAACAAAGCTTTTATCACTCATGTGG 1160
1162 AATGAATGGATCTATGAAGCTATTTACATGAGGCTCCCTATGTTGAGAGTTCCCATAT 1221
1161 AGCCAAATGGATCTATGAAGCTATTTACATGAGGCTCCCTATGTTGAGAGATTCATGTT 1220
1222 TGGATATCAGCTTGAATCAATAGCTCAGATGAAAGCCAAAGAGAGCTGTGAATATA 1281
1221 TGGATATCAGCTTGAATCAATAGCTCAGATGAAAGCCAAAGAGAGCTGTGAATATA 1280
1282 CTTCAAAAATATGACAGGAAAGTTTACTGAGGCTTTGAGACAGTCAATACCGATTC 1341
1281 CTTCAAAAATATGACAGGAAAGTTTACTGAGGCTTTGAGACAGTCAATACCGATTC 1340
1342 CTTCAAAAATATGACAGGAAAGTTTACTGAGGCTTTGAGACAGTCAATACCGATTC 1401
1341 CTTCAAAAATATGACAGGAAAGTTTACTGAGGCTTTGAGACAGTCAATACCGATTC 1400
1402 CTTCAAAAATATGACAGGAAAGTTTACTGAGGCTTTGAGACAGTCAATACCGATTC 1461
1401 CTTCAAAAATATGACAGGAAAGTTTACTGAGGCTTTGAGACAGTCAATACCGATTC 1460
1462 GCGATCAGCTGCGCACTGACCTGTTCCAGACCTAATCTATAGATGATGTTGGTT 1521
1461 TCGAGTTCAGGCGCACTGACCTGTTCCAGACCTAATCTATAGATGATGTTGGTT 1520

1522 CCTGCTGACCTGTGCGCACTGATATATCTTGTTCACAAAATGTTTTATTTCTCTG 1581
1521 CCTGCTGACCTGTGCGCACTGATATATCTTGTTCACAAAATGTTTTATTTCTCTG 1580
1582 TCAAAAATATTAATAAATCTAGAAAAGATAGAAAAGAGGAAATAGATCTTTC 1631
1581 CTGAGAGTTGCTAGAAAAGAGAAAAGAGGAAATAGATTTATATC 1630

RESULT 8
US-09-981-353-189
; Sequence 189, Application US/09981353
; Patent No. US20020160382A1
; GENERAL INFORMATION:
; APPLICANT: Laeak, Amy W.
; APPLICANT: Jones, David A.
; TITLE OF INVENTION: GENES EXPRESSED IN COLON CANCER
; FILE REFERENCE: PA-0038 US
; CURRENT APPLICATION NUMBER: US/09/981,353
; CURRENT FILING DATE: 2001-10-11
; NUMBER OF SEQ ID NOS: 194
; SOFTWARE: PERL Program
; SEQ ID NO 189
; LENGTH: 1712
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID No. US20020160382A1 480489.5
US-09-981-353-189

Query Match 74.7%; Score 1232.4; DB 10; Length 1712;
Best Local Similarity 85.6%; Pred. No. 0;
Matches 1382; Conservative 0; Mismatches 231; Indels 1; Gaps 1;

19 ATTGCATTCATCAGATGTCATGAAATGAGATTCAGCTTCTCTGATACAGCTGAG 78
1 ATCCATTCGACCAAGATGATCTGAAATGAGATTCAGCTTCTCTGATACATCTCAG 60
79 CTGTTACTTTAGCTCTGAGAGTTGTGAAAGTCTGTGTGCTCCACAGAAATTCAGCA 138
61 TTGTTACTTTAGCTCTGAGAGTTGTGAAAGTCTGTGTGCTCCACAGAAATTCAGCA 120
139 CTGATGATTAATAAGCAATCTGATGAACTTGTCCAGAGAGTCAATGAGTGTCTGT 198
121 TTGATGATTAATAAGCAATCTGATGAACTTGTGTGATGAGAGTCAATGAGTGTCTGT 180
199 ATTGCATTCATCAGATGTCATGAAATGAGATTCAGCTTCTCTGATACAGCTGAG 258
181 ACTGCATTCATCAGATGTCATGAAATGAGATTCAGCTTCTCTGATACAGCTGAG 240
259 AGTTATCTGATCTTAACTAAATCTGATGAGATTTATCAAGCAGCTGTTAA 318
241 AGTTATCTGATCTTAACTAAATCTGATGAGATTTATCAAGCAGCTGTTAA 300
319 GAGATGCGCAAGCTTCCAAAGACATTTTGTGATATTTTCAAGATTAAGAAAT 378
301 GAGATGCGCAAGCTTCCAAAGACATTTTGTGATATTTTCAAGATTAAGAAAT 360
379 CATGAGCATTTATATGATCAATCTAGAAAGTCTGTAAGATTAAGTTCATAATAGAA 438
361 CCGTGGGAATTAATATGATCAATCTAGAAAGTCTGTAAGATTAAGTTCATAATAGAA 420
439 ACTATGAGAACTACAGAGTCAAGATTTGATGTTGTTCTTGACAGTCTGTTTCCC 498
421 AGTTATGAAAAAATACAGAGTCAAGATTTGATGTTGTTCTTGACAGTCTGTTTCCC 480
499 CTTTGTGAGCTGCTGCGCGAGTACTTAATAATCCCTTGTCAAGCCTCCGCTTCTC 558
481 CTGTTGTGAGCTGCTGCGCGAGTACTTAATAATCCCTTGTCAAGCCTCCGCTTCTC 540
559 TCTGCTACCAATTTGAAAAAGCATAGTGAAGACTTCTGTTCCCTCTCTATGTCCTC 618

| | | | |
|----|------|--|------|
| Db | 541 | TCCTGGCTACACAAATTGAAGGACAGGTGAGAGACTGATTTTCCCTCCTTCTCAACATACC | 600 |
| OY | 619 | TGTTGTTATGTCAGAACTTAAGTGAACCAATGACTTTCAATAGAGAGGGTAAAAATATGAT | 678 |
| Db | 601 | TATTTGTTATGTCAGAAATTAAAGTGAATCAATGACTTTTCATGAGAGAGGGTAAAAATATGAT | 660 |
| OY | 679 | CTATGTGCTTTATTTTGAATTTTGGTTCCAAAATATTGACATGAAAGTGGGATCGATT | 738 |
| Db | 661 | CTATGTGCTTTATTTTGAATTTTGGTTCCAAAATGTCGTGATATGAAAGTGGGATCGATT | 720 |
| OY | 739 | CTACAGTGAAGTTCTTAGAAGAACCCACTACGTATCTGAGACAAATGGCAAAAGCTGACAT | 798 |
| Db | 721 | TTACAGTGAAGTTTTRAGAAAGACCCACTACCTTATTGAGACAAATGGGAAAGCTGACAT | 780 |
| OY | 799 | ATGGCTTATTCGAAACTACTGAGGATTTTCAATTTTCTCAACCACTTTACCAAAATGTTGA | 858 |
| Db | 781 | ATGGCTTATTCGAAACTCTGGAGTTTCAATTTCTCTACATTCATTTCTCAAAAGTTGA | 840 |
| OY | 859 | GTTTCGTTGGAGACATCCACT - GCAAACTGGCCAAACCCCTTACCGAAGAAATGAAAGAT | 917 |
| Db | 841 | TTTTGTGAGAGATTCCTCAGTGGCAAACTGGCAAAACCCCTTAAAGGAAATGAGAGAT | 900 |
| OY | 918 | TTGTCAGAGCTCTGGAGAAATATGCTGTGGTGTGTTCTCTGGGGGTGATGTCAGTA | 977 |
| Db | 901 | TTGTCAGAGCTCTGGAGAAATATGCTGTGGTGTGTTCTCTGGGGGTGATGTCAGTA | 960 |
| OY | 978 | ACACGTCAGMAAGAAAGGGCCAAATGTAATGCAATCAGCCCTTGGCAAGATCCCAAAAAG | 103 |
| Db | 961 | ACATGACAGCAAGAAAGGGCCAAATGTAATGTCAGACGCCCTTGGCAAGATCCCAAAAAG | 102 |
| OY | 1038 | TTCTGTGAGATTTGATGAGGAAATAAACAGATPACTTTAGAGCTCAATPACTCGGTGTACA | 109 |
| Db | 1021 | TTCTGTGAGATTTTGACGGGAAATAAACAGATGCTTAAAGTCTCAATPACTCGGTGTACA | 108 |
| OY | 1098 | AGTGAATPCCCGAAGATGATCTTCTTGATCACCACCAAGCTTTTATACATCGTAG | 115 |
| Db | 1081 | AGTGAATPCCCGAAGATGATCTTCTTGATCACCACCAAGCTTTTATACATCGTAG | 114 |
| OY | 1158 | GTGGAATGATGGGATCTATGAACTATTACATGGGGTCCCTTATGTTGGAGTTCCCA | 121 |
| Db | 1141 | GTGAGGCCAATGGCATCTATGAGGCAATCTACATGGGATCCCTATGTTGGGATTTCCAT | 120 |
| OY | 1218 | TATTTGGTATCAGCTTGATAACATAGCTCAGATGAAGGCCAAAGAGCAGCTGTAGAAA | 127 |
| Db | 1201 | TGTTTGTGATCAACCTGATTAACATTTGCTCACATGAAGGCCAAAGGAGCAGCTGTAGAT | 126 |
| OY | 1278 | TTAACTTCAAACTATGACAGCGAAGATTACTGAGGGCTTTGAGAACATCATTTACCG | 133 |
| Db | 1261 | TGCACTTCAACCAATGTCGATGACAGACCTGCTGATCACTGAAGACATTAATTAAG | 132 |
| OY | 1338 | ATTTCCTTAATAAGAAATGCTATGAGATTTTCAAGATTCACCATGATCAACCTGTAA | 139 |
| Db | 1321 | ATCCTTTATATAAGAAATATTAAGAAATTTCAAGAAATTCACATGATCAACCTGTAA | 138 |
| OY | 1398 | AGCCCTAGATGAGCAGCTTTCTGGATGAGTTTGTCACTGCGCCAATAAGAGCCCAAC | 145 |
| Db | 1381 | AGCCCTAGATGAGCAGCTTTCTGGATGAGTTTGTCACTGCGCCAATAAGAGCCCAAC | 144 |
| OY | 1458 | ACCTGCGATCAGCTGCCCATGACCTCACTGGTTCCAGACATACCTATATAGATGTGATTG | 151 |
| Db | 1441 | ACCTTGCAGTTCCAGCCCATGACCTCACTGGTTCCAGACATACCTTTGGATGTGATTG | 150 |
| OY | 1518 | GGTTCCTGTCGACCTGTGTGGGAACTGCTAATTTCTGTGTCACAAAATGTTTTTATTTT | 157 |
| Db | 1501 | GGTTCCTGTCGACCTGTGTGGGAACTGCTAATTTTATATCAACAAAATGTTTTGTCTGTTT | 156 |
| OY | 1578 | CCTGTCAAAATTTTATATAAATAGATAGAAAAGAGGAAATGATCTTTTC | 1631 |
| Db | 1561 | GTTTCGAAAGTTTGTCTAGAAAAGGAGAAAGGAAAGATTAATGTTATGTC | 1614 |

```

US-10-158-646-42
; Sequence 42, Application US/10158646
; Publication No. US20030073105A1
GENERAL INFORMATION:
; APPLICANT: Lasek, Amy K.W.
; APPLICANT: Sornasse, Thierry
; TITLE OF INVENTION: GENES EXPRESSED IN COLON CANCER
; FILE REFERENCE: PA-0030-1 US
; CURRENT APPLICATION NUMBER: US/10/158,646
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: 60/295,239
; PRIOR FILING DATE: 2001-05-31
; NUMBER OF SEQ ID NOS: 78
; SOFTWARE: PERL Program
; SEQ ID NO: 42
; LENGTH: 1712
; TYPE: DNA
; ORGANISM: Homo sapiens
FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID No. US20030073105A1 480489.3
US-10-158-646-42

Query Match      74.7%; Score 1232.4; DB 14; Length 1712;
Beet Local Similarity 85.6%; Pred. No. 0;
Matches 1382; Conservative 0; Mismatches 231; Indels 1; Gaps 1

19  ATTCGATTCGACATCAGAGATGCTCATTAATGACATTCAGCTCTTCTCTGATACAGCTGAG 78
Db  1  ATCGGATTCGACACAGAGATGACTCTGAATGACATTCAGCTCTTCTCTGATACATCTCAG 60

79  CTGTACTTTAGCTCTGGAGTTGTGAAAGGTGCTGTGTGTCGCCACAGAAATTCAGCCA 138
Db  61  TTGTTACTTTAGCTCTGGAGTTGTGAAAGGTGCTGTGTGTCGCCACAGAAATTCAGCCA 120

139  CTGATGTAATTAAGAACATTCGGAGTGAACCTTCCAGAGAGTCAATGAGTGAATCT 198
Db  121  TTGATGTAATTAAGAACATTCGAAAGAGCTGTGTGAGAGAGTCAATGAGTGAATCT 180

199  ATTCGATTCCTCAGCTTCCTCATTTCTTGGATCCCAACAGCCCATCTACTTTAAATTTGA 258
Db  181  ACTGGCATCTTCAGCTTCCTCATTTCTTGGATCCCAATGATGATCATCTTTAAATTTGA 240

259  AGTTTATCTGTATCTTTAACTAAACCTGAGTTGAGGATATTAATCAAGCAGCTGTTAA 318
Db  241  AGTTTATCTGTATCTTTAACTAAACCTGAGTTGAGGATATTAATCAATCAGCTGTTAA 300

319  GAGATGGGACGACCTTCCAAAAGACACATTTTGGTCATATTTTTCACAACTCAAGAAAT 378
Db  301  GAGATGGGACGACCTTCCAAAAGAGCTTTGGTTAATTTTTCACAAAGCAAGAAAT 360

379  CATGTCGACATTTAATGACATCTACTAGAAAGTCTGTAAGGATATAGTTCAATTAAGA 438
Db  361  CCGTGGGAATTAATGACATCTTTAGAACTTCTGTAAGGATATAGTTCAATTAAGA 420

439  ACTTATGAAGAACTACAGAGATCAAGATTTGATGTTGCTTTTTCAGATGCTGTTTCCC 498
Db  421  AGTTATGAAGAACTACAGAGATCAAGATTTGATGATCGTTTTCAGATGCTGTTTCCC 480

499  CTTTGTGAGCTGCTGGCCGAGTTACTTAAATACCTTTGTCTACAGCTTCGCTTCTC 558
Db  481  CTTGTCGAGCTGCTGGCCGAGTTACTTAAATACGTTTGTCTACAGCTTCGCTTCTC 540

559  TCCGTCGACGCAATGAAAGACATAGTGGAGGACTTCGTGTTCCCTCCCTCTATGTGC 618
Db  541  TCCGTCGACGCAATGAAAGGACAGTGGAGGACTGATTTTCCCTCTCTCTCATATCC 600

619  TGTGTTATGTACAGAACTAAGTGAACCAATGACTTTCATAGAGAGGTAAATAATGAT 678
Db  601  TATGTTATGTACAAATTAAGTGAATCAATGACTTTCATAGAGAGGTAAATAATGAT 660

679  CTATGTGCTTATTTGAAATTTGGTTCCAAATATTTGACATGAAGAGTGGATCAAGT 738

```

```

FILE REFERENCE: 44921-5028-WO
CURRENT APPLICATION NUMBER: US/09/880.107
CURRENT FILING DATE: 2001-06-14
PRIOR APPLICATION NUMBER: US 60/211,379
PRIOR FILING DATE: 2000-06-14
PRIOR APPLICATION NUMBER: US 60/237,054
PRIOR FILING DATE: 2000-10-02
NUMBER OF SEQ ID NOS: 3950
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 3756
LENGTH: 2799
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: Genbank Accession No. US20020142981a1 X63359
US-09-880-107-3756

Query Match      74.3%  Score 125.2;  DB 10;  Length 2799;
Best Local Similarity 85.6%;  Pred. No. 0;
Matches 1375;  Conservative 0;  Mismatches 228;  Indels 3;  Gaps 1;

26  TGCATCAGATGTCTAAGAAATGAGACTTCTCTCTGATACAGTGAAGCTGTAC 85
    |||||
    2  TGCACAGGAGTGGCTGGAATGAGACTACAG--TTGCGTGATACAACTCAGTTTAC 58

86  TTTAGCTTGGGAGTGTGGAAGGTGTGTGTGGCCACAGAAATTCAGCCACTGATG 145
    |||||
59  TTTAGCTCTGGAGATGTGGGAAAGGTGTGTGTATGGGCGCGAGAAATCAGCTTTGATG 118

146  AATATAAGACAATCCCGAGATGAACTTGCCAGAGAGGTCAATGAGGATCTGTATTGGCA 205
    |||||
119  AATATGAAGACAATCCCGAGAAACCTGTTGAGAGAGGTCAATGAGGATCTGTACTGGCA 178

206  TCTTCAGCTTCATTTCTTTGATCCCAACGCCCATCTACTCTTAAATTGAAATTAT 265
    |||||
179  TCTTCAGCTTCATTTCTTTGATCCCAACGCCCATCTACTCTTAAATTGAAATTAT 238

266  CCTGTATCTTAACTAAAATGAGTTGAGATATTATCAAGCAGCTGTGTTAAGATGG 325
    |||||
239  CCTCATCTTTAACTAAAATGAGTTGAGATATTATCAAGCAGCTGTGTTAAGATGG 298

326  GCAGAACTTCCAAAGACACTTTTGGCATATTTTTCAGAAAGTACAAAGAAATCATGTGG 385
    |||||
299  TCAGAAATTCAAAAGATACATTTTGGTACCTTTTTCAGAAAGTACAAAGAAATCATGTGG 358

386  ACATTAATGACATACCTTGAAGATCTGTGAAGATATAGTTTCAATTAAGAACTTATG 445
    |||||
359  GCAATTAATGACATATTAAGAACTTGTGAAGATATAGTTTCAATTAAGAACTTATG 418

446  AAGAAATCAAGAGTCAAGATTGATGTCTCTGCAAGTGTCTGTTTCCCTTTGGT 505
    |||||
419  AAAAAATCAAGAGTCAAGATTGACATCGTTTTCAGAGATGCTTATTTACCTGTGGT 478

506  GAGCTGCTGGCCGAGTACTTAAATACCTTTGTATCAACACCTCCGCTTCTCCCTGGC 565
    |||||
479  GAGCTGCTGGCCGAGTACTTAAATACCTTTGTATCAACACCTCCGCTTCTCCCTGGC 538

566  TACGCAATTAAGAAAGATGAGAGACTTCTGTTCCTCTCTTATATGCTCTGTTT 625
    |||||
539  TACTCATTTGAAAGCACAGTGGAGATTTATTTTCCCTCTTCTACGTACCTGTTGT 598
    |||||

```

| | | | |
|---|------|---|------|
| Oy | 806 | ATTCGAAACTACTGGGATTTTCAATTTCTCTCAACCCACTTACCAAAATGTTGAGTTGCTT | 865 |
| Db | 779 | ATGCGAAATCTCTGGAATTTTAAATTTTCTCTATCCATTTCTTACAAATGTTGATTTTGTT | 838 |
| Oy | 866 | GGAGAGCTCCACTGCAAACTGCGCAAAACCCCTACCGAAGAAATGGAAGATTTGTCCAG | 925 |
| Db | 839 | GGAGGACTCCACTGCAAACTGCGCAAAACCCCTACCTAAGAAATGGAAGATTTGTACAG | 898 |
| Oy | 926 | AGCTCTGAGAAAATGTTGTTGTGTTTCTCTGGGGTCGATGGTCAATPACACGTCA | 985 |
| Db | 899 | AGCTCTGAGAAAATGTTGTTGTGTTTCTCTGGGGTCATGGTCAATPACATGAC | 958 |
| Oy | 986 | GAAAGAAAGGCCAATGTAATTTGATCAACACCCCTTGCCAAATGCCCAAAAGTTCTGTGG | 1045 |
| Db | 959 | GAAAGAAAGGCCCAAGTAATTTGCAACACCCCTTGCCAAATGCCCAAAAGTTCTTTGG | 1018 |
| Oy | 1046 | AGATTGATGGGAATPAAACAGATCTTTAGGACTCAATPACTCGGCTGTACAGATGATA | 1108 |
| Db | 1019 | AGATTGATGGGAATPAAACAGATCCTTTAGGCTCAATPACTCGACTGTACAGATGATA | 1078 |
| Oy | 1106 | CCCCGAAATGATCTTTCTTGTCATCCCAAAACCAAAAGCTTTTATCACTCATGTGTGAATG | 1165 |
| Db | 1079 | CCCCGAAATGACCTTCTAGTGCATCCAAAAACCAAGCTTTTATPACTCATGTGTGAAC | 1138 |
| Oy | 1166 | AATGGATCTTATGAAGCTATTTTACATGAGGGGTCCTATATGTGTGGAGTTCCCATATTTGCT | 1223 |
| Db | 1139 | AATGGATCTTATGAAGCAATCTACCATGGGATCCCTATATGTGTGGAGTTCCCATATTTGCT | 1198 |
| Oy | 1226 | GATCAGCTTGTATCATATGAGCTCACAATGAAGGCCCAAGGAGCGCTGTATGAATTAACCTTC | 1285 |
| Db | 1199 | GATCAGCTTGTATTAATTTGCTCATATGAAGGCCCAAGGAGCGCTGTATGAGTGAATTC | 1258 |
| Oy | 1286 | AAAATATGACAAAGCAAGGATTTTACTGAGGGCTTTGAGAACAGTCATTACCGATCTCT | 1345 |
| Db | 1259 | AAACAATATGCAAGTACAGACCTGCTGATGATGACAGTAAGACAGTAAATTAATGATCTCTCA | 1318 |
| Oy | 1346 | TATTAAGAGATGCTATGATGATTAATCAAGATTAATCAATGATCAACTGTATTAAGCCCTTA | 1405 |
| Db | 1319 | TATTAAGAGATATATATATAAATTAATCAAGATTAATCAATGATCAACTGTATTAAGCCCTG | 1378 |
| Oy | 1406 | GATGAGAGAGCTTCTGATGATGAGTTTGTATGATGGGCCCAAAAGGAGCAAGACCTGTGA | 1465 |
| Db | 1379 | GATGAGAGAGCTTCTGATGATGAAATTTGTATGATGGGCCCAAAAGGAGCAAAACATTTTGA | 1438 |
| Oy | 1466 | TCAGTGGCCATGACCTCACTGTTCCAGACACTCTCTAATAGATGATTTGGTTCTGTG | 1525 |
| Db | 1439 | GTTGAGAGCCCAACACTCACTCGTTCCAGTACACTCTTTGAGATGATTTGGTTCTGTG | 1498 |
| Oy | 1526 | CTGACCTGTGTGGCACTGCTATATTTCTTTTACAAAATGTTTATTTTCTGTCAA | 1585 |
| Db | 1499 | CTGGCTTGTGTGGCAACCGTGTATTTATATCAAAAGTGTGTCTTTTGTGTTCTGTG | 1558 |
| Oy | 1586 | AAATTTAATTAACCTAGAAAGTATATAAAAGAGGAAATGATCTTTC | 1631 |
| Db | 1559 | AAGTTTGTAGAAAAGAAAGAGGAAAAGGGAATGATTATATC | 1604 |
| RESULT 11 | | | |
| US-09-981-353-45 | | | |
| ; Sequence 45, Application US/09981353 | | | |
| ; Patent No. US20020160382A1 | | | |
| ; GENERAL INFORMATION: | | | |
| ; APPLICANT: Lasek, Amy W. | | | |
| ; APPLICANT: Jones, David A. | | | |
| ; TITLE OF INVENTION: GENES EXPRESSED IN COLON CANCER | | | |
| ; FILE REFERENCE: PA-0038 US | | | |
| ; CURRENT APPLICATION NUMBER: US/09/981.353 | | | |
| ; CURRENT FILING DATE: 2001-10-11 | | | |
| ; NUMBER OF SEQ ID NOS: 194 | | | |
| ; SOFTWARE: PERL Program | | | |
| ; SEQ ID NO 45 | | | |
| ; LENGTH: 2150 | | | |
| ; TYPE: DNA | | | |

| | ORGANISM: Homo sapiens |
|----------------------------|--|
| FEATURE: | |
| NAME/KEY: | misc_feature |
| OTHER INFORMATION: | Incyte ID No. US20020160382A1 255115.4 |
| NAME/KEY: | unSURE |
| LOCATION: | 2087, 2089, 2094, 2096-2098, 2108, 2110, 2112, 2115-2116, 2120, |
| LOCATION: | 2122-2123, 2125, 2136 |
| OTHER INFORMATION: | a, t, c, g, or other |
| US-09-981-353-45 | |
| Query Match | 68.9%; Score 1137.2; DB 10; Length 2150; |
| Best Local Similarity | 82.1%; Pred. No. 6.5e-306; |
| Matches 1345; Conservative | 0; Mismatches 288; Indels 5; Gaps 3 |
| Db | 1 AGCAACTGAAAAACAAGCATTCGATTCGATCAGATGTCATGAATGGAATTCAGCTCT 60 |
| Db | 18 AACAACTGAAAAAGAGCATTCGATTAAGACCGAGATGTCCTGAATGGAATGTCAGCTT 77 |
| Qy | 61 TCTCTGATACAGCTGACTGTACTTTAGCTCTGGAGTTGTGAAAAGTCTGTGTG 120 |
| Db | 78 TCTGTGATGACGCTCACTGTGTAACCTTGGAGTTGTGAAAAGTCTGTGTG 137 |
| Qy | 121 GCCCAGCAAAATTCAGCCCATGGATCAATATAAGCAATCCGATGAGAACTTGCAGAG 180 |
| Db | 138 GCCCAGCAAAATTCAGCCCATGGATTAATATGAGCAAACTCTGGAAGACTTGTCCAGAG 197 |
| Qy | 181 AGGTGATGAGGTGACTGTATTGGACATCTTCAGCTCCATTTCTTTCGATCCCAAGACCC 240 |
| Db | 198 GGGTCATGAGGTGATGTGTGACATCTTCGGCTTCTATTCTGTGCAATGCCAGTAATC 257 |
| Qy | 241 ATCTACTCTTAAATTTGAAGTTTATCCTGATCTTTAACTAAACTGAGTTGAGATAT 300 |
| Db | 258 ATCTGCTTTTAAATTTGAAGTTTATCCATCACTTTAACTAAATTTGAAGATTT 317 |
| Qy | 301 TATCAAGAGCGTGGTAAGATGGGGAGA--ACTGCCAAAGACACTTTGGTCATAT 357 |
| Db | 318 TTTTATGAAAATGTTCGATGAGTGACATATGATATTTCAAAAAATCACTTTGGTCATAT 377 |
| Qy | 358 TTTTTCACAAAGTACAAAGAAATCATGTGACATTTATGACATCTTAGAAAGTCTGTAA 417 |
| Db | 378 TTTTTCACAACTACAAAGAAATGTGTGGGAATATCTGACTATATATTAAGCTCTGTGA 437 |
| Qy | 418 GGATATAGTTCAAAATAAGAACTTTATGAAGAACTACAGAGTCAAGATTGATGTTGT 477 |
| Db | 438 AGATCAGATTTTGAACAAGAAACTTATGAGAAACTACAGAGTCAAAATTTGATGTCT 497 |
| Qy | 478 TCTTGCAAAATCTGTTTTCCCTTTGGAGAGCTGTGCCAGATCACTTAAATATCCCT 537 |
| Db | 498 TCTGCAAAATCTGTTAAATCCCTGTGGAGCTGTGTGCAACTCACTTAAATATCCCT 557 |
| Qy | 538 TGTCTACAGCTCCGCTCTCTCCTGGCTACGCAATTTAAAAAGCATAGTGGAGACTCT 597 |
| Db | 558 TCTGTACAGCTCTCCGCTCTCTGTTGGCTACAGAGTTGAGAAATGAGTGGAGATTTCT 617 |
| Qy | 598 GTTCCCTCTTCTATGTGCTGTTGTTATGTGAGAACTAAGTCAAAATGACTTTTCAAT 657 |
| Db | 618 GTTCCCTCTTCTCTATGTGCTGTTGTTATGTGAGAACTAAGTCAAAATGACTTTTCAAT 677 |
| Qy | 658 AGAGAGGTAATAAATATGATCTATAGTCTTATTTTGAATTTTGGTCCAAATATTTTGA 717 |
| Db | 678 GAGAGGATTAATAAATATGATATATATCTTATTTTGAATTTTGGTTCCAAGCATATGA 737 |
| Qy | 718 CATGAAGAAGTGGAGTCACTTCTACAGTGAAGTTCTAGAAGACCCACTACGTAATCTGA 777 |
| Db | 738 TCTGAAGAAGTGGAGTCACTTCTATATGAGAACTTCTAGGAAGACCCACTACATTTTGA 797 |
| Qy | 778 GACAAATGCAAAAGCTGACATATGAGCTTATTTGAAACTACTGAGATTTTCAATTTCTCA 837 |
| Db | 798 GACAAATGCAAAAGCTGAAATGTGGCTCATTCGAACCTATTGGAGATTTTGAATTTCTCG 857 |
| Qy | 838 CCCACTTTTACCAATGTTGAGTTGTTGGAGAGATCCACTGCAAACTGCGCAAAACCCCT 897 |
| Db | 858 CCCACTTTTACCAATGTTGATTTTGTGGAGAGATTTCACTGTAACAGCGCAAAACCCCT 917 |

```

Qy 898 ACCGAGGAATGGAAGAGTTTGTCCAGAGCTCTGGAGAAATGGTGTGTGTTC 957
Db 918 GCTTAGGAATGGAAGAGTTTGTCCAGAGCTCTGGAGAAATGGTGTGTTC 977
Qy 958 TCTGGGTCGATGTGTGAGTAAACGCTCAGAGAAAGGCGCAATGTAATGTCAGCCCT 1017
Db 978 TCTGGGTCGATGTGTGAGTAAACGCTCAGAGAAAGGCGCAATGTAATGTCAGCCCT 1035
Qy 1018 TGCCAAATCCCAAAAGGTTCTGTGAGATTTATGGAAATTAACCAATCTTGA 1077
Db 1036 TGCCAAATCCCAAAAGGTTCTGTGAGATTTATGGAAATTAACCAATCTTGA 1095
Qy 1078 ACTCAATCTCGGCTGTCAAGTGTATACCCAGAAATGTCCTTGTGATCCCAAG 1137
Db 1096 TTTCAATCTCGGCTGTCAAGTGTATACCCAGAAATGTCCTTGTGATCCCAAG 1155
Qy 1138 CAAGGCTTTTATCACTATGTGTGAATGAATGGATCTATGAAGCTATTTACATGGG 1197
Db 1156 CAAGGCTTTTATCACTATGTGTGAATGAATGGATCTATGAAGCTATTTACATGGG 1215
Qy 1198 CCTATGTGTGAGTTCCTCATTTTGTGTGATGAGCTTGTATACATGCTCAGTAAAG 1257
Db 1216 CCTATGTGTGAGTTCCTCATTTTGTGTGATGAGCTTGTATACATGCTCAGTAAAG 1275
Qy 1258 CAAGGAGAGCTGTGAATAATTAACCTCAAACTATGAAGCAAGCAATTTACTGAGG 1317
Db 1276 CAAGGAGAGCTGTGAATAATTAACCTCAAACTATGAAGCAATTTACTGAGG 1335
Qy 1318 TTTGAGAACAGTCATTAACGATTCCTCTTATTAAGAGATGCTATGAGATTAACAAG 1377
Db 1336 ATTGAAGTCAGTCATTAATGAGCCCTATCTATTAAGAGATTAATGAGATTAACAAG 1395
Qy 1378 TCACATGATCAACTGTGAAGCCCTTGTGATGAGCAGCTTCTGTGATGAGTGTCTAT 1437
Db 1396 TCACATGATCAACTGTGAAGCCCTTGTGATGAGCAGCTTCTGTGATGAGTGTCTAT 1455
Qy 1438 GCGGCAAGAGAGCAAGACCTGCGATCAGCTGCCATGACCTCACCTGGTTCAGCA 1497
Db 1456 GCGGCAAGAGAGCAAGACCTGCGATCAGCTGCCATGACCTCACCTGGTTCAGCA 1515
Qy 1498 CTACTCTATGATGTGATGAGTGTGCTGTGACCTGTGTGAGCACTGCTATATCTGT 1557
Db 1516 CCACCTTTGATGTGATGAGTGTGCTGTGACCTGTGTGAGCACTGCTATATCTGT 1575
Qy 1558 CACAAATGTTTTTATTTTCTGTCAAAATTTAATAAATGAGATGAGAAAG 1617
Db 1576 CACAAATGTTTTTCTGTGTGTTTTTCCGAAGCTGCCAAACAGAGAAAGAAAG 1635
Qy 1618 GGAATAGATCTTTCCAA 1635
Db 1636 GGAATAGATCTTTCCAA 1653

```

```

RESULT 12
US-10-252-157-25
; Sequence 25, Application US/10252157
; Publication No. US20030190640A1
; GENERAL INFORMATION:
; APPLICANT: Paris, Mary
; APPLICANT: Pearson, Cecelia I.
; TITLE OF INVENTION: GENES EXPRESSED IN PROSTATE CANCER
; FILE REFERENCE: PA-0027-1 US
; CURRENT APPLICATION NUMBER: US/10/252,157
; CURRENT FILING DATE: 2002-10-01
; PRIOR APPLICATION NUMBER: 60/295,048
; PRIOR FILING DATE: 2001-05-31
; NUMBER OF SEQ ID NOS: 501
; SOFTWARE: PERL Program
; SEQ ID NO 25
; LENGTH: 2150
; TYPE: DNA
; ORGANISM: Homo sapiens

```

```

; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. US20030190640A1 25115.4
; FEATURE:
; NAME/KEY: unsure
; LOCATION: 2087, 2089, 2094, 2096-2098, 2108, 2110, 2112, 2115-2116, 2120, 2122-2123,
; LOCATION: 2125, 2136
; OTHER INFORMATION: a, t, c, g, or other
US-10-252-157-25

```

```

Query Match 68.9%; Score 1137.2; DB 12; Length 2150;
Best Local Similarity 82.1%; Pred. No. 6.5e-306;
Matches 1345; Conservative 0; Mismatches 288; Indels 5; Gaps 3;

```

```

Qy 1 AGCAATGGAAGAAAGAGATTCATGATGAGATGATGATGATGATGATGATGATGATGAT 60
Db 18 AACAACTGGAAGAAAGAGATTCATGATGAGATGATGATGATGATGATGATGATGATGAT 77
Qy 61 TCTCTGATGAGCTGAGCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 120
Db 78 TCTCTGATGAGCTGAGCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 137
Qy 121 GCCCAAGAAATTCAGCCATGATGATGATGATGATGATGATGATGATGATGATGATG 180
Db 138 GCCCAAGAAATTCAGCCATGATGATGATGATGATGATGATGATGATGATGATGATG 197
Qy 181 AGGTCATGAGGTCATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 240
Db 198 GGGTCATGAGGTCATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 257
Qy 241 ATCTACTCTTAAATTTGAATTTGATTTGATTTGATTTGATTTGATTTGATTTGATTTG 300
Db 258 ATCTACTCTTAAATTTGAATTTGATTTGATTTGATTTGATTTGATTTGATTTGATTTG 317
Qy 301 TATCAAGCAGCTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 357
Db 318 TTTTATGAAATGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 377
Qy 358 TTTTCAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 417
Db 378 TTTTCAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 437
Qy 418 GGAATAGATTCATTAATGAAATCTTAAGAGAAATCAAGAGTCAGATTTGATGAT 477
Db 438 AGATGAGATTTTGAAGAGAAATCTTAAGAGAAATCAAGAGTCAGATTTGATGAT 497
Qy 478 TCTGCAAGATGCTGTTTTTCTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 537
Db 498 TCTGCAAGATGCTGTTTTTCTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 557
Qy 538 TGTCTACAGCTCCGCTTCTCTGCTGCTAGCAATGAAAGCATGTTGAGGACTCT 597
Db 558 TGTCTACAGCTCCGCTTCTCTGCTGCTAGCAATGAAAGCATGTTGAGGACTCT 617
Qy 598 GTTCCCTCTCTCTATGATGCTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 657
Db 618 GTTCCCTCTCTCTATGATGCTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 677
Qy 658 AGAGAGGTTAAATATGATGATGATGATGATGATGATGATGATGATGATGATGAT 717
Db 678 AGAGAGGTTAAATATGATGATGATGATGATGATGATGATGATGATGATGATGAT 737
Qy 718 CATGAAGATGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 777
Db 738 TCTGAAGATGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 797
Qy 778 GACAATGCAAAAGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 837
Db 798 GACAATGCAAAAGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 857
Qy 838 CCACTCTTACCAATGTTGATGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 897
Db 858 CCACTCTTACCAATGTTGATGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 917

```

| | | | |
|--|------|--|------|
| QY | 898 | ACCGAAGAAATGGAAGATTGTCAGAGCTCTGGAAGAAATGGTGTGAGTCTTTTC | 957 |
| Db | 918 | GCTTAGGAATATGGAAGACTTTGTGCAGAGCTCTGGAAGAAATGGTATTGTGTGTTTC | 977 |
| QY | 958 | TCTGGGGTCGATGTCGATTAACACGTCGAGAAGAGGGCCAAATGTAATGCATCAGCCCT | 1017 |
| Db | 978 | TCTGGGGTCGATGATCAGTAAACATGTCAGAGAAGAAAGTGCCAAACATGA-TGCATCAGCCC- | 1035 |
| QY | 1018 | TGCCAAGATCCCAAGAAAGTTCCTGTGAGATTTTATATGGAAATAAACAGATACCTTTAGG | 1077 |
| Db | 1036 | TGCCCAAGATCCCAAGAAAGTTCCTATATGAGATTTATATGCAAGAGCCAAATAACTTTTAGG | 1095 |
| QY | 1078 | ACTCAATACTGGCTGTACAAAGTGATATCCCGAATATCTTTTGGTCATCCCAAAAC | 1137 |
| Db | 1096 | TTCCATATCTGCATCTGTAACAGTGTATCCCGAATATACCTCTTTGGTCATCCCAAAAC | 1155 |
| QY | 1138 | CAAACTTTTATCACTCATGTGTGAATGAATGGATCTATGAAAGTATTTAACATGGGGT | 1197 |
| Db | 1156 | CAAACTTTTATCACTCATGTGTGAACCAATGGCATCTATGAGGCAATCTACATGGGAT | 1215 |
| QY | 1198 | CCCATATGGTGGAGTTCCTCATATTTTGGATATAGCTTGTATTAACATAGCTCATATAGGC | 1257 |
| Db | 1216 | CCCATATGGTGGACATTTCCCTGTGTGGGATCTAACATGATATAACATTTGCTCATAGAAAGC | 1275 |
| QY | 1258 | CAAAAGACAGCTGTAGAAATTAACCTTCAAAACTATGACAAGCGAAGATTTACTAGAGGC | 1317 |
| Db | 1276 | CAAGGACAGACCCCTCAGTGTGCATCATGAGACCATGTCCAGTAGAATTTTGTCTCATAGC | 1335 |
| QY | 1318 | TTTGAAGAACGTCATTACCGATTCTCTTATTAAGAGAAGATGCTATAGATTATCAAGAT | 1377 |
| Db | 1336 | ATTGAAGTCAGTCACTTAATATGAACCCATCTATTAAGAGAATATCATGAATTAATCAAGAT | 1395 |
| QY | 1378 | TCACCATATCAACCTGTAAAGCCCTATGATGAGCACTCTTGGATTCGAGTTTGTAT | 1437 |
| Db | 1396 | TCATCATATCAACCGGTGAAGCCCTGTGATGACAGATCTTGTGATTTGAGTTGTAT | 1455 |
| QY | 1438 | GCGCCACAAGAGGCAAGCAACCTGCGATCAGCTCCCATGATCACTGACTGTTCCAGCA | 1497 |
| Db | 1456 | GCGCCATTAAGAGAGCAAGCAACCTTCCGGGTGCGAGCCCAACAACCTGACTGATTCAGTA | 1515 |
| QY | 1498 | CTACTCTATAGATGTGATTTGGTTCCTGTCGAATAATTTAATAACTAGAAAGATAGAAAAG | 1557 |
| Db | 1516 | CCACTCTTGGATGTGATGATGATTCCTGCTGCGCTGCGCAACTATGATATTTATGAT | 1575 |
| QY | 1558 | CACAAAATGTTTTTATTTCTTCTGTCGAATAATTTAATAACTAGAAAGATAGAAAAG | 1617 |
| Db | 1576 | CACAAAATGTTGCCGTTTTGTTTCGAAAACCTTGCCAAAACAGGAAAGAAAGAAAAG | 1635 |
| QY | 1618 | GGAATAGATCTTTCAAA 1635 | |
| Db | 1636 | GGATTAGTTATATCAAA 1653 | |
| RESULT 13 | | | |
| US-10-252-157-24 | | | |
| : Sequence 24, Application US/10252157 | | | |
| : Publication No. US20030190640A1 | | | |
| GENERAL INFORMATION: | | | |
| : APPLICANT: Paris, Mary | | | |
| : APPLICANT: Pearson, Cecelia I. | | | |
| : TITLE OF INVENTION: GENES EXPRESSED IN PROSTATE CANCER | | | |
| : FILE REFERENCE: PA-0027-1 US | | | |
| : CURRENT APPLICATION NUMBER: US/10/252,157 | | | |
| : PRIOR APPLICATION NUMBER: 60/295, 048 | | | |
| : PRIOR FILING DATE: 2001-05-31 | | | |
| : NUMBER OF SEQ ID NOS: 501 | | | |
| : SOFTWARE: PERL Program | | | |
| : SEQ ID NO 24 | | | |
| : LENGTH: 1829 | | | |
| : TYPE: DNA | | | |
| : ORGANISM: Homo sapiens | | | |

```

; FEATURE:
; NAME:KEY: misc_feature
; OTHER_INFORMATION: Incyte ID No. US20030190640A1 255115.2
US-10-252-157-24

```

Query Match 68.8%; Score 1135.8; DB 12; Length 1829;
Best Local Similarity -81.6%; Pred. No. 1.4e-305;
Matches 1327; Conservative 0; Mismatches 297; Indels 3; Gaps 1,

| | | | |
|----|-----|--|-----|
| QY | 1 | AACAAACATTGCAATGGCAATCAGGAATGCTATGAAATGGAACTTCAGCTTCCTCTGTATAC | 71 |
| Db | 2 | AAGAACATTGGCAATGAACCAAGATGTCTGTAAATGGAACGTCTTTCTGCTGATAC | 61 |
| QY | 72 | AGCTGAGCTGTTACTTACTCTGGAGAGTGTGGAAAGTGTGTGTGCTGCCACAGAT | 131 |
| QY | 192 | TGACTGATTTGGACATCTTCAGCTTCATCTTCTTCGATCCCAAGACCCACTCTACTTA | 251 |
| Db | 182 | TGACTGTGTGACATCTTGGCTTCTACTCTTGTCAATGCCAGTAATCACTGCTATTA | 241 |
| QY | 252 | AATTGAAGTTATCTCTGATCTTTAACTAAACTGAGTTTGAGATATTAACAAGAC | 311 |
| Db | 242 | AATTGAAGTTATCTCTACATCTTTAACTAAAAATGATTTGGAAGATCTCTCTGA | 301 |
| QY | 312 | TGGTTAAGATGGGCA---GACTTCCAAAGACACTTTTGGTCAATTTTTCACAAG | 366 |
| Db | 302 | TTCCTGATGATGATATATATGTGTGTTTAAAAAAATCACTTTTGTGATATTTTTCACAAT | 361 |
| QY | 369 | TACAAGAATTCATGTGACATTTAATGACATCTTAGAAAGTTCTGTAAAGATATAGTTT | 428 |
| Db | 362 | TACAAGAATTTGTTGGGAATTTATATGACTACAGTAACAGCTCTGTAAAGATGCAGTTT | 421 |
| QY | 429 | CAAAATGAACCTTATGAAAGAACTACAGAGTCAGAATTTGATGTTGTTCTTGACAGATG | 488 |
| Db | 422 | TGAATTAAGAACTTATGATGAAACTACAAAGATCAAAAGTTGATGTCAATCTGGCAGATG | 481 |
| QY | 489 | CTGTTTTCCCTTTGGTAGCTGCTGCGCCGAGTTACTTAAATACCCTTTGTCTACAGCC | 548 |
| Db | 482 | CCCTTATCCCTGTGTGTGAGCTACGCTGGAACATTTAATCATCCCTTTCTGTACAGTC | 541 |
| QY | 549 | TCCGCTTCCTCCTGGATACGCAATTTGAAAGCAATGAGAGGACTCTGTCCCTCCTT | 608 |
| Db | 542 | TTGATTTCTCTGTGGCTACACATTTGAAAGAAATGAGAGGATTTCTGTCCCTCTT | 601 |
| QY | 609 | CCATGTGCTCTGTTTATATGTCAGAACTAAGTGAACCAATGACTTTGATAGAGGGTAA | 668 |
| Db | 602 | CCATGTACTCTGTGTTATGTACAGATTTAATGTATCAAAATGATTTTATAGAGAGATTA | 661 |
| QY | 669 | AAAATATGATCTATGTGCTTATTTTGAATTTTGGTTTCCAAATATTTGACATGAAGAT | 728 |
| Db | 662 | AAAATATGATATATATGCTTTATTTTGACTTTTGGTTTCAAGCATATGATCTGAAGAAT | 721 |
| QY | 729 | GGGATCAGTTCAACGTGAAGTCTAAGGAACCCACTACTTATCTGAAGCAATGGCAA | 788 |
| Db | 722 | GGGATCAGTTTATATGTGAAGTCTTAGGGAACCCACTACTTATTTGAGACAAATGGGGA | 781 |
| QY | 789 | AAGCTCACTATAGCTTATTTGAAAACTCTGGGAATTTTCAATTTCTCACCCACTCTTAC | 848 |
| Db | 782 | AAGCTGAATATGGCTCATTTGGAACCTATTTGGGAATTTTGAATTTCTCTGCCCATTTTAC | 841 |
| QY | 849 | CAAAATGTGATGTTGTTGGAGGACTCACTGCAAACTGTGCCAAACCCCTTACGGAAGAAA | 908 |
| Db | 842 | CAAAATGTGATTTTGTGGAGGACTTCACTGTAACCAAGCAAAACCTTGTCTTAAGAAA | 901 |
| QY | 909 | TGGAAAGTTTGTCCAGAGCTCTGGAAGAAATGTGTGTTGTGTGTTTCTCTGGGTGCA | 968 |
| Db | 902 | TGGAAAGTTTGTGAGAGCTCTGGAAGAAATGTGTGTTGTGTGTTTCTCTGGGTGCA | 961 |

Db 965 GAAGAAAGTGCACATGATTTGACATCAGCCCTTGCCCGAGATCCCAAAAGGTTCTATGG 1024
Qy 1046 AGATTTGATGGGAATTAACCGAGATCTTAGAGCTCAATCTCGGCTGTAAAGTGATA 1105
Db 1025 AGATTTGATGGGAATTAACCGAGATCTTAGAGCTCAATCTCGGCTGTAAAGTGATA 1084
Qy 1106 CCCGAGATGATCTTCTGTGATCATCCCAAAACCAAGCTTTATCATCTCATGAGTGAATG 1165
Db 1085 CCCGAGATGATCTTCTGTGATCATCCCAAAACCAAGCTTTATCATCTCATGAGTGAATG 1144
Qy 1166 AATGGATCTATGAGATTAATTTACCATGGGGTCCCTATGATGGAGTTCCCATATTTGGT 1225
Db 1145 AATGGATCTATGAGAGGATCTACCATGGATCCCTATGATGGAGTTCCTGTGTTGGCG 1204
Qy 1226 GATCAGCTGATTAATAGTCTCATCATGAAGCCCAAGAGAGAGCTGTGAATTAATCTTC 1285
Db 1205 GATCAGATGATTAATAGTCTCATCATGAAGCCCAAGAGAGAGCTGTGAATTAATCTTC 1264
Qy 1286 AAAAATATGACAGAGAGATTTACTGAGGGCTTTGAGACAGTCAATTAACGATTCCTCT 1345
Db 1265 AGAAGCATGATGATGAGATTTGCTCATGATGATGATGATGATGATGATGATGATGATGATG 1324
Qy 1346 TATAAGAGAAATGCTATGATGATTAATCAAGAAATTCACATGATCAACCTGTAAAGCCCTA 1405
Db 1325 TATAAGAGAAATGCTATGATGATTAATCAAGAAATTCACATGATCAACCAATGAAGCCCTG 1384
Qy 1406 GATGAGAGATCTTCTGTGATGATGATTTGATGATGATGATGATGATGATGATGATGATG 1465
Db 1385 GATGAGAGATCTTCTGTGATGATGATTTGATGATGATGATGATGATGATGATGATGATG 1444
Qy 1466 TCAGTGGCCCATGACCTGACCGGTTCCAGACATCTATAGATGATGATGATGATGATGATG 1525
Db 1445 GTCGAGCTGACAACTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1504
Qy 1526 CTGACCTGTGTGAGCACTGATTAATCTTGTTCACAAATGTTTTTATTTTCTGTGCA 1585
Db 1505 CTGACCTGTGTGAGCACTGATTAATCTTGTTCACAAATGTTTTTATTTTCTGTGCA 1564
Qy 1586 AAATTTAATTAATAGAGAGATGAGAGAGAGAGAGAGATGATCTTCCAA 1635
Db 1565 AAGCTTGCCAAACAGAGAGAGAGAGAGAGAGATGATTAATTAATCAAA 1614

RESULT 15
US-09-880-107-3292
; Sequence 3292, Application US/09880107
; Patent No. US20020142981A1
; GENERAL INFORMATION:
; APPLICANT: Horne, Darci T.
; APPLICANT: Vockley, Joseph G.
; APPLICANT: Scherif, Uwe
; TITLE OF INVENTION: Gene Expression Profiles in Liver Cancer
; FILE REFERENCE: 44921-5028-WO
; CURRENT APPLICATION NUMBER: US/09/880,107
; CURRENT FILING DATE: 2001-06-14
; PRIOR APPLICATION NUMBER: US 60/211,379
; PRIOR FILING DATE: 2000-06-14
; PRIOR APPLICATION NUMBER: US 60/237,054
; PRIOR FILING DATE: 2000-10-02
; NUMBER OF SEQ ID NOS: 3950
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3292
; LENGTH: 2090
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: Genbank Accession No. US20020142981A1 U08854
US-09-880-107-3292

Query Match 68.2%; Score 1125.2; DB 10; Length 2090;
Best Local Similarity 81.6%; Pred. No. 1.4e-302;

Matches 1314; Conservative 0; Mismatches 293; Indels 3; Gaps 1;
Qy 29 ATGAGATGCTATGAAATGAGATGAGCTCTCTCTCTGATPACAGCTGATGATCTT 88
Db 16 ACCAGATGCTCTGAAATGAGATGAGCTCTCTCTCTGATPACAGCTGATGATCTT 75
Qy 89 AGCTCTGGAGTTGTGAAAGGCTGCTGCTGAGCCACAGAAATTCAGCCATGATGAAT 148
Db 76 AGCTCTGGAGTTGTGAAAGGCTGCTGCTGAGCCACAGAAATTCAGCCATGATGAAT 135
Qy 149 ATAAAGCAATCTGATGAACTTTGTCAGAGAGCTATGAGTGAATGATGATGATGATG 208
Db 136 ATGAAGCAATCTGAGAAAGCTGTTGACAGAGGCTATGAGGATGATGATGATGATGATG 195
Qy 209 TCAGCTTCAATTTCTTTCGATGCCAAGCCCATCTACTGATTAATTTGAATTTATCT 268
Db 196 TCGCTTCTACTCTTGTCAATGCGAGTAATCATCTGCTATTAATTAAGAGTTATCT 255
Qy 269 GATCTTTAATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 328
Db 256 ACATCTTTAATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 315
Qy 329 --GAAGCTTCAAGAGACATTTGTCATATTTTTCACAGTACAGAAATCATGCTG 385
Db 316 TATGCTTTTCAAAATTAATGATGATGATGATGATGATGATGATGATGATGATGATG 375
Qy 386 ACATTTAATGACATCTTGAAGATGATGATGATGATGATGATGATGATGATGATGATG 445
Db 376 GAATTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 435
Qy 446 AAGAACTACAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 505
Db 436 ATGAACTACAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 495
Qy 506 GAGCTGTGCGGAGTACTTAATAATCCCTTGTCTACAGCTCGCTCTCTCTGCTG 565
Db 496 GAGCTGTGCGGAGTACTTAATAATCCCTTGTCTACAGCTCGCTCTCTCTGCTG 555
Qy 566 TACGCAATGAAAGCATGATGAGAGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 625
Db 556 TACGCAATGAAAGCATGATGAGAGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 615
Qy 626 ATGTCAGACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 685
Db 616 ATGTCAGACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 675
Qy 686 CTTTATTTGATTTTGTGTTTCAATATTTGATGATGATGATGATGATGATGATGATGATG 745
Db 676 CTTTATTTGATTTTGTGTTTCAATATTTGATGATGATGATGATGATGATGATGATGATG 735
Qy 746 GAGTTCTAGAGAGACCCATGATGATGATGATGATGATGATGATGATGATGATGATGATG 805
Db 736 GAGTTCTAGAGAGACCCATGATGATGATGATGATGATGATGATGATGATGATGATGATG 795
Qy 806 ATTGAACTAGAGAGATTTTCAATTTCTGACCCATCTTACCAATGTTGATGATGATG 865
Db 796 ATTGAACTAGAGAGATTTTCAATTTCTGACCCATCTTACCAATGTTGATGATGATG 855
Qy 866 GAGAGCTCAGCTGCAAACTGCAAAACCCCTACGAGAGAAATGAGAGATTTGTCAG 925
Db 856 GAGAGCTCAGCTGCAAACTGCAAAACCCCTGCTAAGAAATGAGAGATTTGTCAG 915
Qy 926 AGCTCTGAGAGAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 985
Db 916 AGCTCTGAGAGAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 975
Qy 986 GAAGAAAGGCGCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1045
Db 976 GAAGAAAGTGCACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1035
Qy 1046 AGATTTGATGGGAATTAACCGAGATCTTAGAGCTCAATCTCGGCTGTACAGTGATA 1105
Db 1036 AGATTTGATGGGAATTAACCGAGATCTTAGAGCTCAATCTCGGCTGTACAGTGATA 1095

```
QY 1106 CCCGAGATGATCTTCTGTGTCATCCCAAAACCAAGCTTTTATCACTCATGTGGAATG 1165
    |||||
Db 1096 CCCGAGATGATCTTCTGTGTCATCCCAAAACCAAGCTTTTATCACTCATGTGGAATG 1155
    |||||
QY 1166 AATGGATCTATGAGCTATTTTACCATGGGGTCCCTATGTGGAGTTCCCAATTTGGT 1225
    |||||
Db 1156 AATGGATCTATGAGCGATCTACCATGGGATCCCTATGTGGGATTCCTTGTTCG 1215
    |||||
QY 1226 GATCAGCTTGTATACATAGCTCACTGAGGCAAGAGACAGCTGTAGAAATAACTTC 1285
    |||||
Db 1216 GATCAACATGATTAACATTGCTACATGAAAGCAAGGAGAGCCCTCAGTGGACATC 1275
    |||||
QY 1286 AAACTATGACAAGGAGATTTTACTGAGGGCTTTGAGAACAGTCAATTACGATTCCTCT 1345
    |||||
Db 1276 AGGACATGTCAGATGAGATTTGCTCATATGATGAGTCAATTAATGACCTGTC 1335
    |||||
QY 1346 TATAAGAGAAATGCTATGAGATTAACAAGAAATTCACATGATCAACTGTAAAGCCCTA 1405
    |||||
Db 1336 TATAAGAGAAATGCTATGAGATTAACAAGAAATTCATGATGACCAACCAATGAGCCCTG 1395
    |||||
QY 1406 GATGAGAGATCTTCTGTGATGAGTTTGTCAATGGCCCAAGAGACCAAGCACTGCGA 1465
    |||||
Db 1396 GATGAGAGATCTTCTGTGATGAGTTTGTCAATGGCCCAAGAGACCAAGCACTTGA 1455
    |||||
QY 1466 TCAGTGGCCATGACCTGACCTGGTTCAGCACTACTCTATAGATGATGGGTTCTG 1525
    |||||
Db 1456 GTCGAGCTCAACACTGATCTGATCCAGTACCACTCTTGGATGATAGCATTCCTG 1515
    |||||
QY 1526 CTGACCTGTGTGGCAACTGCTATATCTGTGTCACAAAATGTTTTATTTCTGTCAA 1585
    |||||
Db 1516 CTGGCCTGTGTGGCAACTGCTATATCTGTGTCACAAAATTTGCTGTGTTTCCGA 1575
    |||||
QY 1586 AAATTTAATAAAGTAGAAGATAGAAAAGAGGATAGATCTTCCAA 1635
    |||||
Db 1576 AAGCTTGCAAAACAGAAAGAGAAAGAGATTAGTTATATCAAA 1625
    |||||
```

Search completed: December 7, 2003, 08:52:34
Job time : 414 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 7, 2003, 05:23:08 ; Search time 2466 Seconds
(without alignments)
16262.119 Million cell updates/sec

Title: US-09-980-729B-10
Perfect score: 1650
Sequence: 1 agcaactgcgaagaacagcat.....ccaattcaagaagacctg 1650

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 22781392 segs, 12152238056 residues
Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_estc:*
9: gb_est1:*
10: gb_est2:*
11: gb_est3:*
12: gb_est4:*
13: gb_est5:*
14: em_estfun:*
15: em_estom:*
16: em_esthum:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vrc:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rod:*
26: em_gss_phg:*
27: em_gss_vrl:*
28: gb_gss1:*
29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|----------|-------------|
| 1 | 961.8 | 58.3 | 1946 | AK050435 | Mus muscu |
| 2 | 955.4 | 57.9 | 2573 | AK004971 | Mus muscu |
| 3 | 890.8 | 54.0 | 2895 | AK083294 | Mus muscu |
| 4 | 886.8 | 53.7 | 1892 | AK034801 | Mus muscu |

| | | | | | |
|----|-------|------|------|-----------|-----------|
| 5 | 864.2 | 52.4 | 1896 | AK002736 | Mus muscu |
| 6 | 785.8 | 47.6 | 1783 | AK050327 | Mus muscu |
| 7 | 781 | 47.3 | 2099 | AK008601 | Mus muscu |
| 8 | 769 | 46.6 | 2575 | BC048920 | Mus muscu |
| 9 | 606.8 | 36.8 | 983 | BX444042 | BX444042 |
| 10 | 600.6 | 36.4 | 960 | BQ713091 | AGENCOURT |
| 11 | 599.8 | 36.4 | 1004 | BQ942104 | AGENCOURT |
| 12 | 580.8 | 35.2 | 595 | BG562901 | 602581752 |
| 13 | 576.4 | 34.9 | 927 | BG428781 | 602500860 |
| 14 | 575.4 | 34.9 | 927 | BQ925596 | AGENCOURT |
| 15 | 573.4 | 34.8 | 588 | BSM087819 | AGENCOURT |
| 16 | 560 | 33.9 | 560 | CB153390 | K-EST0210 |
| 17 | 546.4 | 33.1 | 558 | CB113345 | K-EST0155 |
| 18 | 545.8 | 33.1 | 559 | BSM087698 | Homo sapi |
| 19 | 545.8 | 33.1 | 697 | BG4277192 | 602494236 |
| 20 | 543.8 | 33.0 | 560 | BSM087740 | Homo sapi |
| 21 | 529.8 | 32.1 | 823 | BFB689099 | 602185172 |
| 22 | 527.2 | 32.0 | 930 | BQ934372 | AGENCOURT |
| 23 | 523.6 | 31.7 | 977 | B1311895 | 602982580 |
| 24 | 522.8 | 31.7 | 785 | BG429706 | 602493711 |
| 25 | 515.6 | 31.2 | 891 | BG189133 | RST8173 A |
| 26 | 511.8 | 31.0 | 662 | CB148234 | K-EST0204 |
| 27 | 511.6 | 31.0 | 770 | BG201991 | RST21337 |
| 28 | 510.8 | 31.0 | 748 | BG431126 | 602498753 |
| 29 | 506.8 | 30.7 | 510 | BSM073698 | BSM073698 |
| 30 | 496.4 | 30.1 | 1030 | BQ714919 | Homo sapi |
| 31 | 492.6 | 29.9 | 746 | BFB68309 | 602185172 |
| 32 | 488.2 | 29.6 | 787 | BG430305 | 602502309 |
| 33 | 485.4 | 29.4 | 901 | BG400539 | 602464778 |
| 34 | 484.4 | 29.4 | 756 | BG429801 | 602494817 |
| 35 | 483.8 | 29.3 | 728 | BG433410 | 602501120 |
| 36 | 480.8 | 29.1 | 913 | BQ714368 | AGENCOURT |
| 37 | 477.8 | 29.0 | 801 | BG212458 | RST32045 |
| 38 | 471.6 | 28.6 | 1005 | BG167171 | 602344696 |
| 39 | 465.8 | 28.2 | 618 | CB113377 | K-EST0155 |
| 40 | 460.6 | 27.9 | 692 | CB154247 | K-EST0212 |
| 41 | 458.8 | 27.8 | 759 | B1330877 | 602981265 |
| 42 | 456.8 | 27.7 | 821 | BG402643 | 602465575 |
| 43 | 450.2 | 27.3 | 871 | BQ900474 | AGENCOURT |
| 44 | 443 | 26.8 | 923 | A1529976 | u187g04.y |
| 45 | 441.4 | 26.8 | 938 | BQ942682 | AGENCOURT |

ALIGNMENTS

RESULT 1
AK050435
LOCUS
DEFINITION Mus musculus adult male liver tumor cDNA, RIKEN full-length
enriched library, clone: C730047M05
Product: UDP-GLUCURONOSYLTRANSFERASE 2B1 PRECURSOR, MICROSMAL (EC 2.4.1.17) (UDPgt) (UDPgtR-2) homolog [Rattus norvegicus], full insert sequence.
ACCESSION AK050435
VERSION AK050435.1 GI:26341183
KEYWORDS HTC; CAP trapper.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
REFERENCE
1 Carninci, P. and Hayashizaki, Y.
High-efficiency full-length cDNA cloning
Meth. Enzymol. 303, 19-44 (1999)
99279253
PUBMED 10349636
REFERENCE
2 Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Kono, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes
Genome Res. 10 (10), 1617-1630 (2000)

| | | |
|---|---|--|
| MEDLINE PUBMED REFERENCE AUTHORS | 20499374 11042159 3 | Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P., Kono, H., Akiyama, J., Nishi, K., Kitenana, T., Tashiro, H., Itoh, M., Suni, N., Ishii, Y., Nakamura, S., Hazama, M., Nishimori, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kasahigashi, K., Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watabiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y. |
| TITLE | RIKEN integrated sequence analysis (RISA) system-384-format sequencing pipeline with 384 multicapillary sequencer | |
| JOURNAL MEDLINE PUBMED AUTHORS | Genome Res. 10 (11), 1757-1771 (2000) 20530913 11076861 4 | Kawai, J., Shinagawa, A., Shibata, K., Yoshino, M., Itoh, M., Ishii, Y., Arakawa, T., Hara, A., Fukunishi, Y., Kono, H., Adachi, J., Fukuda, S., Aizawa, K., Izawa, M., Nishi, K., Kiyosawa, H., Kondo, S., Yamataka, I., Saito, T., Okazaki, Y., Gojobori, T., Bono, H., Kasukawa, T., Saito, R., Kadota, K., Matsuda, H., Ashburner, M., Batalov, S., Casavant, T., Plechmann, W., Gaasterland, T., Gissi, C., King, B., Kochiwa, H., Kuehl, P., Lewis, S., Matsuo, Y., Nikaide, T., Pesci, G., Quackenbush, J., Schriml, L. M., Staudl, F., Suzuki, R., Tomita, M., Wagner, L., Washio, T., Sakai, K., Okido, T., Furuno, M., Aono, H., Baldarelli, R., Barsh, G., Blake, J., Boffelli, D., Bojunga, N., Carninci, P., de Bonaldo, M. F., Brownstein, M. J., Bult, C., Fletcher, C., Fujita, M., Gariboldi, M., Gietzsch, S., Hill, D., Hotmann, M., Hume, D. A., Kamiya, M., Lee, N. H., Lyons, P., Marchionni, L., Mashima, J., Mazzaretti, P., Mombauts, P., Nordone, P., Ring, B., Ringwald, M., Rodriguez, I., Sakamoto, N., Sasaki, H., Sato, K., Schonbach, C., Seva, T., Shibata, Y., Storch, K. F., Suzuki, H., Toyokata, K., Wang, K. H., Weitz, C., Whitaker, C., Wilmberg, L., Wyshak-Boris, A., Yoshida, K., Hasegawa, Y., Kawaji, H., Kohsaki, S. and Hayashizaki, Y. |
| TITLE | Functional annotation of a full-length mouse cDNA collection | |
| JOURNAL MEDLINE PUBMED AUTHORS | Nature 409 (6821), 685-690 (2001) 21085660 11217851 5 | The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team. Analyses of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs Nature 420, 563-573 (2002) |
| JOURNAL AUTHORS | 6 (bases 1 to 1946) | Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P., Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, M., Hayashida, K., Hayatsu, N., Hiramoto, K., Hiroaka, T., Hirozane, T., Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T., Kato, H., Kawai, J., Kojima, Y., Kondo, S., Kono, H., Kouda, M., Koye, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M., Nakamura, M., Nishi, K., Nomura, K., Nunazaki, R., Ohno, M., Ohsato, N., Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S., Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A., Muramatsu, M. and Hayashizaki, Y. |
| TITLE | Direct Submission | |
| JOURNAL | Submitted (16-JUN-2001) | Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-res@gs.riken.go.jp, URL: http://genome-gsc.riken.go.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216) |
| COMMENT | | CDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in Riken Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. Tissue was provided by William A. Held, Roswell Park Cancer Institute, Department of Molecular and Cellular Biology, Elm and |

Carlton Streets, Buffalo, NY 14263, whose assistance we gratefully acknowledge.
Please visit our web site for further details.
URL: <http://genome-gsc.riken.go.jp/>
URL: <http://fantom.gsc.riken.go.jp/>.
Location/Qualifiers

FEATURES

source

CDS

```

1..1946
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="PANTOM,DB:C730047M05"
/db_xref="taxon:10090"
/clone="C730047M05"
/sex="male"
/cell_type="tumor"
/tissue_type="liver"
/clone_id="RIKEN full-length enriched mouse cDNA library"
/dev_stage="adult"
34..1623
/note="unnamed protein product;
UDP-glucuronosyltransferase 2B1 precursor, microsomal (EC
2.4.1.17) (UDPGT) (UDPGR-2) homolog [Rattus norvegicus]
(SWISSPROT P09875, evidence: PASTY, 85.8%ID, 100%length,
match=1587)
putative"
/codon_start=1
/protein_id="BAC34254.1"
/db_xref="GI:26341184"
/translation="MSMKQASVFLIPICTIRGACGKLVMPREYSHINMKIILD
ELVRGADVIVLSSASILIGPSNENISINELISAPLSKDDLEAFKVGWNTYEIK
LPLMTISYSLQKISSEYSDMIESFCFAVNMNLSMKLGDSKFDVLAALVPGEL
LSELKLPVLYLRFQCYCEKSGCLPSPYVPLVLELSHMTFAEYKVMGLV
LLEPFWOTENKSNQFSDVLRPTTLTVMGKADIMLVRTFMDLKPFPFLPND
FVGLHCKPKPLPKMEFVOSGEGVVFVFSVMVKNKEKAVVASALAOIQ
KVMRFEGKPDITGSNTRLYKMPDLDLGHPTKAFIANGCTNGYEALYHPIPV
GIPLEGQPDNINII VAKGAIVRVDFVMSITDILAKTVINDPSIKENMRSLRH
HDQPMKPLDRAVFWIEVYMRNKGKHLRPLALDLTWQYSLDVIIGFLCVAVVFL
IAKCLFPCFKTANMGKKKE"
BASE COUNT      564 a      405 c      420 g      557 t
ORIGIN
Query Match      58.3%; Score 961.8; DB 11; Length 1946;
Best Local Similarity 75.4%; Pred. No. 3.2e-192;
Matches 1210; Conservative 0; Mismatches 392; Indels 3; Gaps 1;

```

```

QY      13 ACAAGCATGATGATGCATGAGATGCTATGAATGACCTTCCCTCGATCA 72
DB      12 AAAAACAATTCCTTCAGCAAGATGCTATGAACAGCTTCAGTTTCTGTTGATCA 71
QY      73 GCTGAGCTGTTACTCTTGGCTGTGGAGTGTGGAAGGCTGTGTCGCCACAGAAAT 132
DB      72 GTTCATATGCTATATTTAGCTGAGCCTGTGGAAAGTGCTGTGTCCTACAGATA 131
QY      133 CAGCAGCATGATGAATTAAGACATCTCGATGAACCTTCCAGAGAGCTCATGAGT 192
DB      132 CAGCAGCATGATGAATTAAGACATCTCGATGAACCTTCCAGAGAGCTCATGAGT 191
QY      193 GACGTATGTCATCTTCAGCTTCATTTCTTGATGCCAAGCCATCATCTTAA 252
DB      192 CACGTTCTCAATCTTCTGCTTCATCTTCATTCATTCGCGCGCAATGAATCTTATTA 251
QY      253 ATTGAAGTTATCTGATCTTAACTAAAGTGTGAGATATTTATCAAGCACT 312
DB      252 TTTGAAATTTATCTGACCTTTGAGTAAAGATGATCTGAATGTTTGAAGAAATG 311
QY      313 GGTAAAGATGGCA--GAATCCAAAAGACATTTGTGATATTTTCAAGT 369
DB      312 GTTAGAATCTGACATCGAATTAAGAAATCTTCATTTGACATCTTATTAAGT 371
QY      370 ACAAGAAATCATGTGACATTTAATGACATCTTGAAGAGTTCTGTAAGATATAGTTTC 429
DB      372 GCAAAAATCTCCAGTGAATATTCAGACATGATTAAGAAAGTTCTGCAAGAGATGTTTG 431

```

QY 430 AAATAGAACTTATGAAAGAACTACAGAGCTCAAGATTGATGTCTTTCAGAGATGC 489
 Db 432 GAACAAGAGCTCATGAAAAAACTCAAGAGATCTAAGTTTATGTCTTACAGAGATGC 491
 QY 490 TGTTCCTCCCTTGTGTGAGCTGCTGCGGAGTACTTAAATACCCTTTGTCTACAGCTT 549
 Db 492 CTGTGTTCCCTGTGTGAGCTGCTATCAGAACTGCTTAAAGACACCTTTAGTATACAGTCT 551
 QY 550 CCGCTTCCTGCTGGAGCTGCAATTGAAAGCATATGAGAGAGACTGCTCCCTCCCTC 609
 Db 552 CCGCTTCCTGCTGGAGCTGCAATTGAAAGCATATGAGAGAGACTGCTCCCTCCCTC 611
 QY 610 CTATGTGCTGCTGTTTATATGTCAGAACTAAGTACCAATGACTTTCATAGAGAGGTTAA 669
 Db 612 CTATGTGCTGCTGTTTATATGTCAGAACTAAGTACCAATGACTTTCATAGAGAGGTTAA 671
 QY 670 AAATATGATCTATGTGCTTTATTTTGAATTTTGGTCCAAATATTGACATGAAGAAATG 729
 Db 672 GAAATATGTCAGAGTGTGCTTTTGAATTTTGGTCCAAATATTGACATGAAGAAATCCTG 731
 QY 730 GGATCAGTTCTCAGTGAAGTTCATAGAGAGCCAGTACCTTATCTGAGCAATGGCAAA 789
 Db 732 GAAATCAGTTTCAAGTGAAGTTCATAGAGAGCCAGTACCTTATCTGAGATGAGGAAA 791
 QY 790 AGTGAACATATGCTTATTCGAAACTACTGGGATTTTCAATTTCTCACCCTCTTAAC 849
 Db 792 GCGAGACATATGCTTATTCGAAACTACTGGGATTTTCAATTTCTCACCCTCTTAAC 851
 QY 850 AAATGTTGAGTTCGTTGAGAGACTTCACCTGCAAACTGCGCAAACTCCCTACCGAAGAAAT 909
 Db 852 TAAATTTGACCTTTGTTGAGAGACTTCACCTGCAAACTGCGCAAACTCCCTACCGAAGAAAT 911
 QY 910 GGAAGAGCTTGTCCAGAGCTCTGAGAGAAATGCTGTTGTGTTTCTCTGGGGTGAT 969
 Db 912 GGAAGAAATTTGTGAGAGCTCTGAGAGAAATGCTGTTGTGTTTCTCTGGGGTGAT 971
 QY 970 GGTGAGTACACAGTCAAGAGAAAGGCAATGTAATTCATCAGCCCTTGGCAAGATCCC 1029
 Db 972 GGTAAAAACATTAAGAGAGAAAGGCAATGTAATTCATCAGCCCTTGGCAAGATCCC 1031
 QY 1030 ACAAAGGTTCTGTGAGAGATTTGATGGGAATTAACAGATTAATTGAGACTCAATCTAC 1089
 Db 1032 ACAGAGAGTTCTGTGAGAGATTTGATGGGAATTAACAGATTAATTGAGACTCAATCTAC 1091
 QY 1090 GCTGTACAGAGTATACCCCAAGATGATCTTCTGTGTATCTCCCAAAACCAAGCTTTTAT 1149
 Db 1092 GCTGTACAGAGTATACCCCAAGATGATCTTCTGTGTATCTCCCAAAACCAAGCTTTTAT 1151
 QY 1150 CACTCATGTGGAATGGAATGGATCTATAGAGCTATTACCATGGGGTCCCTATGTGGG 1209
 Db 1152 AGCTCATGTGGAATGGAATGGATCTATAGAGCTATTACCATGGGGTCCCTATGTGGG 1211
 QY 1210 AGTTCCATATTTGTGTGATCAGCTTGTATACAGTCACTGAGAGCCCAAGAGCAGC 1269
 Db 1212 TATTCCTGTTGTTGGGATCAACCTGATTAATTAACACATAGAGCAAGGAGGAGC 1271
 QY 1270 TGTGGAATTAACCTCAAACTATGACAGAGCAAGATTTACTGAGGGCTTTGAGAACAT 1329
 Db 1272 TGTGAGTGTGAGCTTGTATCAATGTCAATACAGAGACCTTCTCAGCTGCTGAAGACTGT 1331
 QY 1330 CATTAACGATTCCTCTTAAGAGAAATGATAGATTAATCAAGAAATTCACCATGATCA 1389
 Db 1332 CATTAACGATTCCTCTTAAGAGAAATGATAGATTAATCAAGAAATTCACCATGATCA 1391
 QY 1390 ACCTGTAAAGCCCTAGATCAGAGCTTCTGTGATGAGTGTGTGATGCGCCCAAGAG 1449
 Db 1392 GCCATGAGAGCCCTTGTGAGCGAGGCTTCTGTGATGAGTGTGTGATGCGCCCAAGAG 1451
 QY 1450 AGCCAGAGACCTGGATGATAGCTGCCCATGACCTCACTGTTCCAGACATACCTTAAGA 1509
 Db 1452 AGCCAGAGACCTGGATGATAGCTGCCCATGACCTCACTGTTCCAGACATACCTCTGGA 1511
 QY 1510 TGTGATTTGGGTTCTGCTGAGACCTGTGTGGCAAGCTGATATATTCTTGTTCACAAATGTTT 1569

Db 1512 TGTGATTTGGGTTCTTATTTGCTGTGTGTGAGCTGTGATTCATTCGAAAAATGTTG 1571
 QY 1570 TTTATTTTCTGTCAAAAATTTAATTAATAACTAGAAAAATAGAAA 1614
 Db 1572 CCTCTTTTGTGCTAAGAGCTGCTAACATGAGAAAGAAAA 1616
 RESULT 2
 AK004971
 LOCUS
 DEFINITION
 Mus musculus adult male liver cDNA, RIKEN full-length enriched library, clone:1300012D20 product:UDP-GLUCURONOSYLTRANSFERASE 2B1 PRECURSOR, MICROSMAL (EC 2.4.1.17) (UDPGr) (UDPGr-2) homolog [status nonregius], full insert sequence.
 ACCESSION
 AK004971
 VERSION
 AK004971.1 GI:12836559
 KEYWORDS
 HTC; CAP trapper.
 SOURCE
 Mus musculus (house mouse)
 ORGANISM
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1
 Carninci, P. and Hayashizaki, Y.
 High-efficiency full-length cDNA cloning
 Meth. Enzymol. 303, 19-44 (1999)
 99279253
 10349636
 2
 Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Kono, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
 Normalization and subraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes
 Genome Res. 10 (10), 1617-1630 (2000)
 20499374
 11042159
 3
 Shibata, K., Itoh, M., Aizawa, K., Nagaoke, S., Sasaki, N., Carninci, P., Kono, H., Akiyama, J., Nishi, K., Kitanai, T., Tashiro, H., Itoh, M., Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watanishi, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsumura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.
 RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer
 Genome Res. 10 (11), 1757-1771 (2000)
 20530913
 11076861
 4
 Kawai, J., Shingawa, A., Shibata, K., Yoshino, M., Itoh, M., Ishii, Y., Arakawa, T., Hara, A., Fukunishi, Y., Kono, H., Adachi, J., Fukuda, S., Aizawa, K., Izawa, M., Nishi, K., Kiyosawa, H., Kondo, S., Yamataka, I., Saito, T., Okazaki, Y., Gojobori, T., Bono, H., Kasukawa, T., Saito, R., Kadoya, K., Matsuda, H., Ashburner, M., Batalov, S., Casavant, T., Fleischmann, W., Gaasterland, T., Gissi, C., King, B., Kochiwa, H., Kuenl, P., Lewis, S., Matsuo, Y., Nikaido, I., Peeole, G., Quackenbush, J., Schriml, L.M., Staib, F., Suzuki, R., Tomita, M., Wagner, L., Washio, T., Sakai, K., Okido, T., Furuno, M., Aono, H., Baldarelli, R., Barsh, G., Blake, J., Boffelli, D., Bojunga, N., Carninci, P., de Bonaldo, M.F., Brownstein, M.J., Bull, C., Fletcher, C., Fujita, M., Gariboldi, M., Guerinich, S., Hill, D., Hofmann, M., Hume, D.A., Kamiya, M., Lee, N.H., Lyons, P., Marchionni, L., Mashima, J., Mazzarelli, J., Mombetta, P., Nordone, P., Rung, B., Ringwald, M., Rodriguez, I., Sakamoto, N., Sasaki, H., Sato, K., Schonbach, C., Seya, T., Shibata, Y., Storch, K.F., Suzuki, H., Toyooka, K., Wang, K.H., Weitz, C., Whitaker, C., Wilmberg, L., Wyszynski, A., Yoshida, K., Hasegawa, Y., Kawai, H., Kohzuki, S. and Hayashizaki, Y.
 Functional annotation of a full-length mouse cDNA collection
 Nature 409 (6821), 685-690 (2001)
 21085660
 11217851
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 AUTHORS

REFERENCE
AUTHORS

5 The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.
Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs
Nature 420, 563-573 (2002)
6 (bases 1 to 2573)

JOURNAL
AUTHORS

Adechi,J., Mizawa,K., Akahira,S., Akiyama,T., Ara,A., Aono,H., Arakawa,T., Bono,H., Caminici,P., Fukuda,S., Furukoshi,K., Furuno,M., Hanagaki,T., Hara,A., Hayatsu,N., Hiramoto,K., Hirooka,T., Hori,F., Imotani,K., Ishii,Y., Itoh,M., Izawa,M., Kasukawa,T., Kato,H., Kawai,J., Kohji,Y., Konno,H., Kouda,M., Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Nishi,K., Nomura,K., Nunataki,R., Ohno,M., Okazaki,Y., Okido,T., Owa,C., Saito,H., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki,D., Shibata,K., Shibata,Y., Shinnagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F., Tanaka,T., Tejima,Y., Toyota,T., Yamamura,T., Yasunishi,A., Yoshida,K., Yoshino,M., Muramatsu,M. and Hayashizaki,Y.

TITLE
JOURNAL

Direct Submission
Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute: 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan [E-mail:genome-resesec.riken.go.jp, URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222, Fax:81-45-503-9216]

COMMENT

Please visit our web site (<http://genome.gsc.riken.go.jp/>) for further details.

CDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. First strand cDNA was primed with a primer [5' GAGGAGAGACGGCGGCACACTCAGATTTTCTTTTTTTTNN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. Second strand cDNA was prepared with the primer acceptor of sequence[5' GAGGAGAGAGATCCCAAGTCATTATTAATTAACAACCCCCCCC 3']. cDNA was cleaved with XhoI and SacI. Cloning sites, 5' end: SacI; 3' end: XhoI. Host: SOLR.

FEATURES
source

Location/Qualifiers

1..2573

/organism="Mus musculus"

/mol_type="mRNA"

/strain="C57BL/6J"

/db_xref="FANTOM DB:1300012D20"

/db_xref="MGI:1904432"

/db_xref="taxon:10090"

/clone="1300012D20"

/sex="male"

/tissue.type="liver"

/clone.lib="RIKEN full-length enriched mouse cdna library"

/dev_stage="adult"

1..2573

/note="UDP-GLUCURONOSYLTRANSFERASE 2B1 PRECURSOR, MICROSMAL (EC 2.4.1.17) (UDPGr) (UDPGr-2) homolog [Rattus norvegicus] (SWISSPROT|P09875, evidence: PASTY, 85.8%ID 100%length, match=1587)"

/db_xref="MGI:1919023"

BASE COUNT 733 a 541 c 497 g 802 t

ORIGIN

Query Match 57.9%; Score 955.4; DB 11; Length 2573;
Best Local Similarity 75.1%; Pred. No. 7.3e-191;
Matches 1206; Conservative 0; Mismatch 396; Indels 3; Gaps 1

Dp 11 AAAAATTCTCTTCAGCAAGATGTCTATGAAGAAGCTTCTCTGATACA 72
13 ACAAGATTGATTCATCATGAGATGCTATGAATGAACTTGAGCTTCTCTGATACA 72
11 AAAAATTCTCTTCAGCAAGATGTCTATGAAGAAGCTTCTCTGATACA 70
73 GCTAGCTGTACTTCTAGCTCTGGAGCTTGGAAAAGTCTGCTGTGGCCACAGAAAT 132

| | | | |
|----|------|---|------|
| Dp | 71 | GTTCATATGCTAATTATAGACTGGAGCCTGAGGAAAGTGTGGTGGCTTACAGATA | 130 |
| Qy | 133 | CAGCCACTGATGAAATATAAAGCAATCTGATGAACCTGTCCAGAGAGTCAATGAGT | 192 |
| Dp | 131 | CAGCCATTGGATAAATATGAAAAAATATCTGGAGAACTGTCCAGAGAGTCAATGAGT | 190 |
| Qy | 193 | GACTGATATGGCATCTTCAGCTTCATCTTCTTGATGCCAACAGCCATCTACTCTTAA | 252 |
| Dp | 191 | CACCGTCTCATCTCTCTGCTTCATCTCTATTTGGGCCAGAGCAATGAATCTTATTTA | 250 |
| Qy | 253 | ATTGGAATTTATCTGTATCTTTAACTAAACCTGATTTGAGATATATCAAGAGCT | 312 |
| Dp | 251 | TTTGAATTTATTTCTGACCTTTGAGTAAAGATGATCTGAATATGCTTTTGAATAATG | 310 |
| Qy | 313 | GATTAAAGATGGGCA---GACTTCCAAAAGACATTTTGGTCATATTTTACAGT | 369 |
| Dp | 311 | GGTGAAGAACTGGGACATACGATTTAAAAAACTTCCATTTTGGACATCTTAATTAACCT | 370 |
| Qy | 370 | ACAGAATATCATGAGACATTTATATGACATCTAGAAAGTTCTGTAAGATATAGTTTC | 429 |
| Dp | 371 | GCAAAAAATCTCCAGTGAATTTAGACATGATTTGAAAGTTTCTGCAAGCAGTATGTTG | 430 |
| Qy | 430 | AAATAGAACTTTATGAAAGAACTACAGAGTCAAGATTTGATTTGTTCTTGCAGATGC | 489 |
| Dp | 431 | GACACAGAGCCTCATGAAAAAACTCCAAGATCTAAGTTGATGTGCTGTCAGACATGC | 490 |
| Qy | 490 | TGTTTTCCCTTTGGTGAAGTCTGCGCGAGTACTTAAATACCCTTTGTCTACAGCT | 549 |
| Dp | 491 | CTTGATTCCTCTGTGTGAGTGTGCTATCAAGACTCTTAAACACTTTTGTATCAATCT | 550 |
| Qy | 550 | CCGCTTCCTCTGTGAGTACGCAATTGAAAGCATATGAGAGAGCTTGTGTCCTCTC | 609 |
| Dp | 551 | CCGCTTCCTCTGCGATATCTATGTAAGAAATCAAGTGGGGCCTTTCACCTCCCTTC | 610 |
| Qy | 610 | CTATGCGCTTTGTATATGTCAGAACTTAAGTGAACCAATGACTTTCATAGAGGGTAAA | 669 |
| Dp | 611 | CAATGTGCTGTGCTCTGTCAAGACTTAAGTGAACCAATGACTTTCATAGAGGGTAAA | 670 |
| Qy | 670 | AAATATGATCTATGTCTTATTTTGAATTTTGGTCCAAATATTTGACATGAAGAACTG | 729 |
| Dp | 671 | GAATATGTTGAGAGTGTGCTTTTGAATTTTGGTCCAAATTTAACGAATCTCTG | 730 |
| Qy | 730 | GGATCAGTTCTACAGTGAAGTTCTTAGAAGACCACTACGTTATCTAGACAAATGGGCAA | 789 |
| Dp | 731 | GAATCAGTTCTACAGTGAAGTTCTTAGAAGACCTTACCAATTTACTGATGATGGGGA | 790 |
| Qy | 790 | AGCTGACATATGAGCTTATTTGAAACTACTGGAGTTTTCATTTTCTCACCCACTTTACC | 849 |
| Dp | 791 | GGCAGACATATGAGCTCTTTCGACACTCTTGGGACTTGAAATTTCTCACCTTTCTTGCC | 850 |
| Qy | 850 | AAATGTGATGTTGTTGAGGAGCTCCACTGCAAACTTGCMAACCCCTTACCGAAGAAAT | 909 |
| Dp | 851 | TAAATTTGACTTTGTGTGAGAGACTCCATTTGAAACCAAGCCAAACCACTGCTTAAGAAAT | 910 |
| Qy | 910 | GGAAGATTTTGTCCAGAGCTCTGGAGAAATATGTGTGTGTGTGTCTCTGGGGTGGAT | 969 |
| Dp | 911 | GGAAGAAATTTTGTTCAGAGCTCTGGAGAAACATGTGTGTGTGTGTCTCTGGGATCAAT | 970 |
| Qy | 970 | GSTCAGTAAACGTTCAGAAAGAAAGGACCAATGTATTTGATCATCAGCCCTTGCAGATACC | 1029 |
| Dp | 971 | GSTTAAACATTTAAAGAAAGAAAGGACCAATGTATTTGATCTCTGTCTTGCAGATACC | 1030 |
| Qy | 1030 | ACAAAGGTTCTGTGAGATTTGATGGAAATAAACAGATACCTTTAGAGCTCAATACTCG | 1089 |
| Dp | 1031 | ACAGAGGTTCTGTGAGATTTGATGGTAAAGAAACAGACACTTTAGGATCCAACACTCG | 1090 |
| Qy | 1090 | GCTGTACAGTGAATCCCAAGATGATCTTCTTGGTCACTCCCAAAACCAAGCTTTTAT | 1149 |
| Dp | 1091 | GCTGTACAGTGAATCCCAAGATGATCTTCTTGGTCACTCCCAAAACCAAGCTTTTAT | 1150 |
| Qy | 1150 | CATCATGTGTGAATGATGGAGTCTATGAAGCTATTACCATGGGGGCTCTATGTTGGG | 1209 |
| Dp | 1151 | AGCTATGTGTGAACCAATGGGATCTATGAAGGCGATCTTCAATGGCATTTCTTATTTGGG | 1210 |

QY 1210 AGTCCATATTTGGTGTGATGCTTGTATACATAGCTCAGATGAGGCCAAGGAGCAGC 1269
 |||||
 Db 1211 TATTCCTTGTGGTGGATGATCACTGATATATTAACCACTAGTACCGAAGGAGCAGC 1270
 |||||
 QY 1270 TGTGAAATTAATCTTCAAACTATGACAGGAAAGATTACTGAGGGCTTTGAGACAGT 1329
 |||||
 Db 1271 TGTTAGAGTGAAGCTTGTATACATATGTCATACAGACCTTCTCAGCTGGTGAAGACTGT 1330
 |||||
 QY 1330 CATTAACGATTCCTCTTATATAGAGAAATGCTATAGATATACAGATTCACCATGATCA 1389
 |||||
 Db 1331 CATTATATGACCTTCTCTATATAAGAAAGCCATAGATTAATCCAGAAATCCACCATGACCA 1390
 |||||
 QY 1390 ACTGTAAAGCCCTAGATGAGCAGTCTTCTGATGAGTTGTGATGAGCCCAAGG 1449
 |||||
 Db 1391 GCCAATGAGCCCTTGAACCGACGCTGCTTTGATGAGATGATGTCATGCGCAACAAAGG 1450
 |||||
 QY 1450 AGCCAGACCTCTGAGATCAGCTGCCCATGACCTGATGTCAGACATCTATAGA 1509
 |||||
 Db 1451 AGCCAGACCTCTGAGATCAGCTGCCCATGACCTGATGTCAGACATCTATAGA 1510
 |||||
 QY 1510 TGTGATTTGGTCTCTGCTGACCTGTGTGGCACTGCTATATCTTTGTCACAAATGTTT 1569
 |||||
 Db 1511 TGTGATTTGGTCTCTGCTGATGCTGTGTGTGATGCTGTATCTATCTGCAAAATGTTG 1570
 |||||
 QY 1570 TTTATTTTCTGTCAAAATTTAATAAAGTAAAGTAGAGAA 1614
 |||||
 Db 1571 CCTTTTGTGTCATAGACTGCTACATGAGGAAAGAGAGAA 1615
 |||||

RESULT 3
 AK083294
 LOCUS
 DEFINITION Mus musculus adult male liver tumor CDNA, RIKEN full-length
 enriched library, clone: C730031G17 product: similar to
 UDP-GLUCURONOSYLTRANSFERASE 2B5 PRECURSOR, MICROSMAL (EC 2.4.1.17)
 (UDPgt) (M-1) [Mus musculus], full insert sequence.
 AK083294.1 GI:26350414
 ACCESSION AK083294.1 GI:26350414
 VERSION
 KEYWORDS HTC; CAP trapper.
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus

REFERENCE
 AUTHORS Carninci, P. and Hayashizaki, Y.
 TITLE High-efficiency full-length cDNA cloning
 JOURNAL Meth. Enzymol. 303, 19-44 (1999)
 MEDLINE 99279253
 PUBMED 10349636

REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 AUTHORS

3
 Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P.,
 Konno, H., Akiyama, J., Nishi, K., Kitsuuni, T., Tashiro, H., Itoh, M.,
 Sumi, N., Ishii, Y., Nakamura, S., Hazama, S., Nishibe, T., Harada, A.,
 Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ilegami, T., Kashiwagi, K.,
 Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Obara, E., Watanabe, M.,
 Okazaki, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J.,
 Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A., and Hayashizaki, Y.
 RIKEN integrated sequence analysis (RISA) system-384-format
 sequencing pipeline with 384 multicapillary sequencer
 Genome Res. 10 (11), 1757-1771 (2000)

Arakawa, T., Hara, A., Fukunishi, Y., Konno, H., Adachi, J., Fukuda, S.,
 Aizawa, K., Izawa, M., Nishi, K., Kiyosawa, H., Kondo, S., Yamataka, I.,
 Saito, T., Okazaki, Y., Gojohori, T., Bono, H., Kasukawa, T., Saito, R.,
 Kadota, K., Matsumoto, H., Ashburner, M., Batalov, S., Casavola, T.,
 Fleischmann, W., Gaasterland, T., Gissi, C., King, B., Kochiwa, H.,
 Kiehl, P., Lewis, S., Matsuo, Y., Nikaido, I., Pesole, G.,
 Quackenbush, J., Schiraldi, L. M., Staudt, F., Suzuki, R., Tomita, M.,
 Wagner, L., Mashio, T., Sakai, K., Okido, T., Furuno, M., Aono, H.,
 Baldarelli, R., Barish, G., Blake, J., Boffelli, D., Bojunga, N.,
 Carninci, P., de Bona, M. F., Brownstein, M. J., Bult, C.,
 Fletcher, C., Fujita, M., Gariboldi, M., Guelinich, S., Hill, D.,
 Hofmann, M., Hume, D. A., Kamiya, M., Lee, N. H., Lyons, P.,
 Marchionni, L., Mashima, J., Mazzarelli, J., Mombauts, P., Nordone, P.,
 Ring, B., Ringwald, C., Rodriguez, I., Sakamoto, N., Sasaki, H.,
 Sato, K., Schonbach, C., Sessa, T., Shibata, Y., Storch, K. F., Suzuki, H.,
 Toyooka, K., Wang, K. H., Weitz, C., Whitaker, C., Wilting, L.,
 Wynshaw-Boris, A., Yoshida, K., Hasegawa, Y., Kawai, H., Kohseki, S.,
 and Hayashizaki, Y.
 Functional annotation of a full-length mouse cDNA collection
 Nature 409 (6821), 685-690 (2001)
 21085660
 MEDLINE
 PUBMED
 REFERENCE
 AUTHORS

5
 The FANTOM Consortium and the RIKEN Genome Exploration Research
 Group Phase I & II Team.
 Analysis of the mouse transcriptome based on functional annotation
 of 60,770 full-length cDNAs
 Nature 420, 563-573 (2002)
 6 (bases 1 to 2895)
 Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P.,
 Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W.,
 Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T.,
 Horii, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T.,
 Kato, H., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Kouda, M.,
 Koya, S., Kuhiwara, C., Matsuyama, T., Miyazaki, A., Murata, M.,
 Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ono, M., Ohsato, N.,
 Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N.,
 Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T.,
 Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S.,
 Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A.,
 Muramatsu, M., and Hayashizaki, Y.
 Direct Submission
 Submitted (16-APR-2002) Yoshihide Hayashizaki, The Institute of
 Physical and Chemical Research (RIKEN), Laboratory for Genome
 Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
 RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
 Kanagawa 230-0045, Japan (E-mail: genome-res@isc.riken.go.jp,
 URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222,
 Fax: 81-45-503-9216)
 CDNA library was prepared and sequenced in Mouse Genome
 Encyclopedia Project of Genome Exploration Research Group in Riken
 Genomic Sciences Center and Genome Science Laboratory in Riken.
 Division of Experimental Animal Research in Riken contributed to
 prepare mouse tissues.
 Tissue was provided by William A. Held, Roswell Park Cancer
 Institute, Department of Molecular and Cellular Biology, Elm and
 Carlton Streets, Buffalo, NY 14263, whose assistance we gratefully
 acknowledge.
 Please visit our web site for further details.
 URL: http://genome.gsc.riken.go.jp/
 URL: http://fantom.gsc.riken.go.jp/
 location/Qualifiers

1. 2895
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="C57BL/6J"
 /db_xref="FANTOM_DB:C730031G17"
 /db_xref="taxon:10090"
 /clone="C730031G17"
 /sex="male"
 /cell_type="tumor"
 /tissue_type="liver"
 /clone_lib="RIKEN full-length enriched mouse cDNA library"

FEATURES
 source
 Kawai, J., Shinagawa, A., Shibata, K., Yoshino, M., Itoh, M., Ishii, Y.,

CDS

```
/dev stage="adult"
33..1632
/note="unnamed protein product: putative
similar to UDP-GLUCURONOSYLTRANSFERASE 225 PRECURSOR,
MICROSOMAL (EC 2.4.1.17) (UDPgt) (M-1) [mus musculus]
(SwissProt|P17117, evidence: FASTA, 82.6%id, 100%length,
match=1587)"
/codon_start=1
/protein_id="BAC38847.1"
/db_xref="GI:26350415"
/translation="MPVKMISALLLLQNSCFRSTSCGKLVLPFESHMNLKIID
ELVORGEVTVLRPSASIFVDPKYSFGIKFETPTAPSKDYLETFLKLVDEMFEV
RDTCYSGLQTLFGKLSYSLCKEASVKQMTKQESKPVLLIDMAASGCELI
AOLIOPELVSRSPGVYOVKNSGGFLPSYVPLVLSGGQMDTEFRYKMICL
YFDTRQPTFEKENDOPFSETIGRPTLLJEMGKAEWAFISYMDLEPRPTLPLNVEY
VGGHCKPAPLPKEMEDFVOSGSHGVVSLSNMNSMTBERANLSMALDIPK
VLWRFDEKTPASLGNTRIYKMLPONDLLHPKTKAPFTHGANGLEYAHHGIPMG
IPLFSEQHDNIHNVAKTAVANLRITMSRDLNLALEVINNSYKENVMLSTIIR
DQMKPLNRTIFMIEFVHRHKGAKHLRPLAHLNLTWYQVHSLDVLGVATAVLVP
VKCLFVYGFVKKKRKNK"
```

BASE COUNT 824 a 566 c 607 g 898 t
ORIGIN

```
Query Match 54.0%; Score 890.8; DB 11; Length 2895;  
Best Local Similarity 72.3%; Pred. No. 3e-177;  
Matches 1156; Conservative 0; Mismatches 442; Indels 0; Gaps 0;
```

| | | | |
|----|-----|--|-----|
| QY | 31 | CAGATGCTCTAATGAAATGACCTTCTCTCTGATACAGCTGACTGATTTAG | 90 |
| DB | 29 | CAGATGCTCTGAAAGTGAATTTCTGCTCTCTCTGCAAGATGATGCTGCTAG | 88 |
| QY | 91 | CTCTGGAGTTGGAAGAGTCTGCTGAGGCCACAGAAATTCAGCACTGATGAT | 150 |
| DB | 89 | ATCCACCGAGCTGGGAAGGTTTGTGTGCGCCACGTGAATTCATGCTGATTAAT | 148 |
| QY | 151 | AAAGCAATCTGTGATGAACTTGTCCAGAGAGTCAAGAGTCACTGATGATGAT | 210 |
| DB | 149 | AAAGTAATATCTGATGAACTTGTCCAGAGAGTCAAGAGTCACTGATGATGAT | 208 |
| QY | 211 | AGCTTCCATTTCTTTCATCCCAACAGCCCATCTCTTAATTTGAAGTTATCTGT | 270 |
| DB | 209 | AGCTTCCATTTCTTTCATCCCAACAGCCCATCTCTTAATTTGAAGTTATCTGT | 268 |
| QY | 271 | ATCTTAATTAATCAAGTGTGAGATATTATCAAGCAAGCTGTTAAGATGGGACA | 330 |
| DB | 269 | AGCTTCAATTAATCAAGTGTGAGATATTATCAAGCAAGCTGTTAAGATGGGACA | 328 |
| QY | 331 | ACTTCCAAAGACATTTTGTGATATTTTTCACAAAGTCAAGAAATCATGTGACAT | 390 |
| DB | 329 | TGAGGTCCCAAGATATCTGTAATCTTCTGCTTCAACACAGCTGTTGTAATTT | 388 |
| QY | 391 | TAAATGATCTTAAAGAGTCTGTAGAGATATAGTTCAATTAAGAACTTATGAAGA | 450 |
| DB | 389 | ATCTGATTAATCTTAAAGTCTGTAGAGATATAGTTCAATTAAGAACTTATGAAGA | 448 |
| QY | 451 | ACTAGAGAGTCAAGATTTGATGTTCTTCTGAGATGCTGTTCCCTTGTGAGGT | 510 |
| DB | 449 | ATTCCAGAGATCAAGTTGATGTTCTTCTGAGATGCTGTTCCCTTGTGAGGT | 508 |
| QY | 511 | GCTGGCCGAGTACTTAAATAACCTTTGTCTACAGCTCCGCTTCTCTGAGTACGC | 570 |
| DB | 509 | CATAGCTCAACGCTCCAGATCTTCTTCTGACAGTCTTGTGCTCTCTGAGTACCA | 568 |
| QY | 571 | AATTAAGAAAGCATAGTGAAGACTTCTGTTCTCTTATGAGCTGTTGTTATGTC | 630 |
| DB | 569 | AGTTAAAAAGAAAGAGTGAAGACTTCTTCTCTTATGAGCTGTTGTTATGTC | 628 |
| QY | 631 | AGAATAGTGAAGAAAGTGAAGACTTGAAGAGGAAAGAAATGATGATGATGCTTA | 690 |
| DB | 629 | AGGATTAAGTGAAGAAAGTGAAGACTTGAAGAGGAAAGAAATGATGATGATGCTTA | 688 |
| QY | 691 | TTTGAATTTTGGTCCAAATATTTGATGATGAAGAGGAGTCACTTCAAGTGAAGT | 750 |

| | | | |
|----|------|--|------|
| DB | 689 | TTTTGATCTTTGGTTCAGACATTTTACAGAAAGAAATGGATATCTTATCACTGAAC | 748 |
| QY | 751 | TCATAGAAAGCCACATGATTTATCTGAGCAATGGCAAAAGTGAATATGCTTATTCG | 810 |
| DB | 749 | TTTGGGAAGGCCACACCTTATATGAGCAATGGGCAAAAGTGAATATGCTTATTCG | 808 |
| QY | 811 | AAATCTGAGATTTTCAATTTTCTCCACCACTTTACCAAAATGTTGATGCTTGAAG | 870 |
| DB | 809 | ATCTCACTGGGATCTGAGATTTCTTCTCACCACTTTACCAAAATGTTGATGATGAG | 868 |
| QY | 871 | ACTTCATCTGCAAACTTCCCAAACTTCCCAAGAGAAATGAGAGTTTCTCCAGACTC | 930 |
| DB | 869 | ACTTCATCTGCAAACTTCCCAAACTTCCCAAGAGAAATGAGAGTTTCTCCAGACTC | 928 |
| QY | 931 | TGAGAAATAGTGTGAGTTTCTCTGAGGATGATGATGATGATGATGATGATGATG | 990 |
| DB | 929 | TGAGATCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG | 988 |
| QY | 991 | AAGGCCAATGTAATTCATCAGCCCTTGCACAAATCCCAAAAGTTCTGTGAGATT | 1050 |
| DB | 989 | AAGGCCAATGTAATTCATCAGCCCTTGCACAAATCCCAAAAGTTCTGTGAGATT | 1048 |
| QY | 1051 | TGATGGCAATTAACCAATCTTATGAGATCTCAATCTGCTGATGATGATGATGATG | 1110 |
| DB | 1049 | TGATGGCAATTAACCAATCTTATGAGATCTCAATCTGCTGATGATGATGATGATG | 1108 |
| QY | 1111 | GAATGATCTTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG | 1170 |
| DB | 1109 | GAATGATCTTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG | 1168 |
| QY | 1171 | GATCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG | 1230 |
| DB | 1169 | ACTCTATGAGCAATTCATGATGATGATGATGATGATGATGATGATGATGATGATG | 1228 |
| QY | 1231 | GCTTGAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA | 1290 |
| DB | 1229 | ACATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA | 1288 |
| QY | 1291 | TATGACAAAGCAAGATTTTCTGAGGCTTGTGAGAAAGTCAATCAATCTCTTATA | 1350 |
| DB | 1289 | AATGCAAGTCAATTTTCTGAGGCTTGTGAGAAAGTCAATCAATCTCTTATA | 1348 |
| QY | 1351 | AGAGATGCTATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT | 1410 |
| DB | 1349 | AGAGATGCTATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT | 1408 |
| QY | 1411 | AGCATCTTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG | 1470 |
| DB | 1409 | GACTATATTTTGAATGATGATGATGATGATGATGATGATGATGATGATGATGATG | 1468 |
| QY | 1471 | TGCCATGACCTCACTGATGATGATGATGATGATGATGATGATGATGATGATGATG | 1530 |
| DB | 1469 | TGCCATGACCTCACTGATGATGATGATGATGATGATGATGATGATGATGATGATG | 1528 |
| QY | 1531 | CTGTGTGCAACTGATTAATTTCTGTTCAAAAATGTTTTTATTTTCTCTGCAAAAT | 1590 |
| DB | 1529 | CTGTGTGTGCAACTGATTAATTTCTGTTCAAAAATGTTTTTATTTTCTCTGCAAA | 1588 |
| QY | 1591 | TAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT | 1628 |
| DB | 1589 | TAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT | 1626 |

RESULT 4
AK034801 1892 bp mRNA linear HTC 05-DEC-2002
LOCUS
DEFINITION
Mus musculus 12 days embryonic body between diaphragm region
and neck CDNA, RIKEN full-length enriched library, clone:9430941C03
product:UDP-glucuronosyltransferase 2 family, member 5, full insert
sequence.
ACCESSION
AK034801 GI:26330209
VERSION
KEYWORDS
HTC; CAP trapper.

SOURCE
ORGANISM
Mus musculus (house mouse)

REFERENCE
AUTHORS
Carninci, P. and Hayashizaki, Y.
TITLE
High-efficiency full-length cDNA cloning
JOURNAL
Meth. Enzymol. 303, 19-44 (1999)
MEDLINE
99279253
PUBMED
10349636

REFERENCE
AUTHORS
Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Kono, H., Okazaki, Y., Muramatsu, M., and Hayashizaki, Y.
TITLE
Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes
JOURNAL
Genome Res. 10 (10), 1617-1630 (2000)
MEDLINE
20499374
PUBMED
11042159

REFERENCE
AUTHORS
Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P., Kono, H., Akiyama, J., Nishi, K., Kusunagi, T., Tashiro, H., Itoh, M., Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A., Yamamoto, R., Watanabe, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watabiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A., and Hayashizaki, Y.
TITLE
RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer
JOURNAL
Genome Res. 10 (11), 1757-1771 (2000)
MEDLINE
20530913
PUBMED
11076861

TITLE
Kawai, J., Shingawa, A., Shibata, K., Yoshino, M., Itoh, M., Ishii, Y., Aikawa, T., Hara, A., Fukunishi, Y., Kono, H., Adachi, J., Fukuda, S., Aizawa, K., Izawa, M., Nishi, K., Kiyosawa, H., Kondo, S., Yamana, I., Saito, T., Okazaki, Y., Gojibori, T., Bono, H., Kasukawa, T., Saito, R., Kadota, K., Matsuda, H., Ashburner, M., Batalov, S., Casavant, T., Fleischmann, W., Gaasterland, T., Gissi, C., King, B., Kochiwa, H., Kuehl, L., Lewis, S., Matsuo, Y., Nikaido, I., Pesole, G., Quackenbush, J., Schriml, L. M., Staudt, F., Suzuki, R., Tomita, M., Wagner, L., Washio, T., Sakai, K., Okido, T., Furuno, M., Aono, H., Baldarelli, R., Barsh, G., Blake, J., Boffelli, D., Bojunga, N., Carninci, P., de Bonaldo, M. F., Brownstein, M. J., Bull, C., Fletcher, C., Fujita, M., Gariboldi, M., Gustincich, S., Hill, D., Hofmann, M., Hume, D. A., Kamiya, M., Lee, N. H., Lyons, P., Marchionni, U., Mashima, J., Mazzarelli, J., Mombaerts, P., Nordone, P., Ring, B., Ringwald, M., Rodriguez, I., Sakamoto, N., Sasaki, H., Sato, K., Schonbach, C., Seya, T., Shibata, Y., Storch, K. F., Suzuki, H., Toyokawa, K., Wang, K. H., Weitz, C., Whitaker, C., Wilming, L., Wynshaw-Boris, A., Yoshida, K., Hasegawa, Y., Kawaji, H., Kohetsuki, S., and Hayashizaki, Y.
TITLE
Functional annotation of a full-length mouse cDNA collection
JOURNAL
Nature 409 (6821), 685-690 (2001)
MEDLINE
21085660
PUBMED
11217851

REFERENCE
AUTHORS
The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.
TITLE
Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs
JOURNAL
Nature 420, 563-573 (2002)
MEDLINE
12118927
PUBMED
12118927

ADACHI, J., Aizawa, K., Akimura, T., Aikawa, T., Bono, H., Carninci, P., Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W., Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T., Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T., Kato, H., Kawai, J., Kojima, Y., Kondo, S., Kono, H., Kouda, M., Koya, S., Kuibara, C., Matsuyama, T., Miyazaki, R., Murata, M., Nakamura, M., Nishi, K., Nomura, K., Numaizaki, R., Ohno, M., Ohnato, N., Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Shibata, K., Shingawa, A., Shiraki, T., Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S., Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A.,

Muramatsu, M. and Hayashizaki, Y.
TITLE
Direct Submission
JOURNAL
Submitted (16-Jul-2001) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-res@sc.riken.go.jp, URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216)
COMMENT
cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in Riken. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.
Please visit our web site for further details.
URL: http://genome.gsc.riken.go.jp/
URL: http://fantom.gsc.riken.go.jp/
Location/Qualifiers
1. 1892
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="FANTOM_DB:9430041C03"
/db_xref="taxon:10090"
/clone="9430041C03"
/tissue_type="embryonic body between diaphragm region and neck"
/clone_1lb="RIKEN full-length enriched mouse cDNA library"
/dev_stage="12 days embryo"
34. 1626
/note="unnamed protein product;
UDP-glucuronosyltransferase 2 family, member 5
(MGI:98900)
putative"
/codon_start=1
/protein_id="BAC28835.1"
/db_xref="GI:26330210"
/translation="MPGKMSIALLLQISCPFSYKGVLYWPMFESHMKIKITLD
ELVORHETVYRPSAYVYLDPKRSGLGKPEPVSVDNLENPIKRDWTYMP
RDTCLSYSLQMDIDBSDFYSLKRDVSNKELMTKQESFVLDLSPVASCBL
LLELLQIPLYLSIRFSPQYIEKSSGFLPLPSYVYVILSGLOMTFLERVNMICR
LVEDFWFOHFNKRMDSFYSEYGRPTTLAETMFKAKEMMLIRSNMDEPFLPVPD
VYGLGHLCKRAKPLPKMDEEFOVSSGDGVVFLSGSVNMTSEKNTAMALAOIPQ
KYLKREDGKTPPLTGKTRVYKMLPONDILGHFKTAPVTHGGANGVEVYIYIGIMI
GIPLEBQNDNTAHVAKGAATVNTIRMSRSVYALALSEVIDNPYKNAITLSTIH
HDDPIPLRAVFWFVFRHRAKRLRSLGHLTWQYHFLDVLFLSCAVTVL
TVKCLFLYRFPVKEKKIKNE"

BASE COUNT
563 a 356 c 393 g 580 t
ORIGIN

Query Match 53.7%; Score 886.8; DB 11; Length 1892;
Best Local Similarity 72.5%; Pred. No. 2.1e-176;
Matches 1162; Conservative 0; Mismatches 437; Indels 3; Gaps 1;

30 TCAGATGCTAAGTAAGTGAAGTCTGCTCTTCCGATACAGCTGAGCTGTAATCTTA 89
29 TCGATGCTGGAAGTGAAGTGAAGTCTGCTCTGCTGCTGCGATGATGATGCTGCTCA 88
90 GCTCTGGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 149
89 GATCTGGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 148
150 TAAAGCAATCTGTAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 209
149 TAAAGCAATCTGTAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 208
210 CAGCTTCGATTTCTTTCGATTCGATTCGATTCGATTCGATTCGATTCGATTCGATTCGATTC 269
209 CAGCTTCGATTTCTTTCGATTCGATTCGATTCGATTCGATTCGATTCGATTCGATTCGATTC 268
270 TATCTTAACTAAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 327
269 CATCTGCTGTAAGCAATCTGTAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 328

328 -AGACTTCCAAAGACATTTTGTGATATTTTTCACAGTACAGAAATCATGTGGA 386
 329 ATGAAATGCCAAGGATACATGTTTGTATTTCTCTTACTGCAAAATATGATGATG 388
 387 CATTAAATGACATCTAGAAAGTCTGTAGAGATATAGTTTCAATTAAGAACTTATGA 446
 389 AATTTTGTGATTTTCTAAGTCTTTTAAAGATGTTGTTTCAAAACAAAGAGCTCATGA 448
 447 AGAAATCTAGAGAGTCAAGATTTGATGTTGTTGCAAGATGCTTTTCCCTTTGGTG 506
 449 CAAAGCTACAGAGATCCAAAGTTGATGCTTTGTGCAATCCTGTCCTTCTGTGGGG 508
 507 AGCTGTCGGGAGGATTAATAAATACCTTTGTCTACAGCTCCGCTTCTCTGAGCT 566
 509 AGCTCATAGCTGAACTGCTCCAGATTTCTTTCTGTACAGATCCGCTTCTCTCAGGCT 568
 567 ACGCAATTAAGAAAGCATAGTGAAGAGACTTGTCTCTCTCTATGTCCTGTGTTA 626
 569 ACCAAATGAAAGTCCAGCGGAGATTTCTACTCCCTCTTATGTAACCTGTAATTC 628
 627 TGTCAAGCTAAGTACCAATGACTTTCTATAGAGAGGTAAATATGATCTATGTC 686
 629 TGTCAAGATTAAGTGAACAAATGACATTCATAGAGAGATAAATATGATGACAGGC 688
 687 TTTATTTTGAATTTTGTTCCTCAATATTTGACATGAAGAGGGATCAGTTCTACAGTG 746
 689 TTTATTTGACTTTTGTTCCTCAATATTTGATTAAGAAATGGACATCTTTTACAGTG 748
 747 AAGTTCTAGAAAGACCACTACCTTATCTGAGACATGGAAGCAAAAGCTGACATATGCTTA 806
 749 AATATTTGGGAAGGCCACACCTTACCTGAGACATGGAAGCAAAATGTTGCTCA 808
 807 TTGCAATCTACTGGGATTTTCAATTTCTCTACCCACTCTTACCAATGTTGAGTTGCTTG 866
 809 TTAGATCCAACTGGGATTTGAGTTTCTTCAACCACTTACCAATGTTGACATATGTTG 868
 867 GAGGCTCCACCTGCAAACTGCAAAACCTTACCGAAGGAAGTGAAGGTTTGCAGA 926
 869 GAGGCTCCACCTGCAAACTGCTTAACTTTGCTTAAAGATATGGAATTTTGTCCAGA 928
 927 GCTCTGAGAAATGATGTTGTGTTGTTCTCTGAGGTCATGTTGATGATGATGATGATG 986
 929 GCTCTGAGAACATGTTGTGTTGTTCTCTGAGGTCATGTTGATGATGATGATGATG 988
 987 AAGAAAGGCCATATTAATGATGATGATGATGATGATGATGATGATGATGATGATG 1046
 989 AAGAAAGGCCAATTAATGATGATGATGATGATGATGATGATGATGATGATGATG 1048
 1047 GATTGATGGGAAATTAACAGATCTTTAGAGACTCAATCTGAGCTGATCAAGATGATG 1106
 1049 AATTGATGGGAAATTAACAGATCTTTAGAGACTCAATCTGAGCTGATCAAGATGATG 1108
 1107 CCCAGATGATCTTCTGCTGATCCCAAAACCAAGCTTTTATCACTGATGATGATG 1166
 1109 CCCAGATGATCTTCTGCTGATCCCAAAACCAAGCTTTTATCACTGATGATGATG 1168
 1167 ATGGAATCTATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1226
 1169 ATGGAATCTATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1228
 1227 ATGGAATCTATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1286
 1229 AATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1288
 1287 AACTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1346
 1289 AACTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1348
 1347 AATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1406
 1349 AATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1408

QY 1407 ATGAGCAGTCTTCTGATGATGATGATGATGATGATGATGATGATGATGATG 1466
 Db 1409 ACAGGGCTGCTTCTGATGATGATGATGATGATGATGATGATGATGATGATG 1468
 QY 1467 CAGCTGCCATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1526
 Db 1469 CACTGGACATTAACCTTACCTGACACGATGATGATGATGATGATGATGATG 1528
 QY 1527 TGACCTGTGGCACTGATGATGATGATGATGATGATGATGATGATGATGATG 1586
 Db 1529 TCTCTGTGTGACATGATGATGATGATGATGATGATGATGATGATGATGATG 1588
 QY 1587 AATTAATAAAGTGAAGATGATGATGATGATGATGATGATGATGATGATG 1628
 Db 1589 TCTTTGAAAGAAAGAAAGAAATTAAGATGATGATGATGATGATGATG 1630
 RESULT 5
 AK002736
 LOCUS
 DEFINITION
 Mus musculus adult male kidney cDNA, RIKEN full-length enriched library, clone:0610033E06 product:UDP-glucuronosyltransferase 2 family, member 5, full insert sequence.
 ACCESSION
 AK002736
 VERSION
 AK002736.1 GI:12832933
 KEYWORDS
 HTC; CAP trapper.
 SOURCE
 Mus musculus (house mouse)
 ORGANISM
 Mus musculus
 Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 REFERENCE
 1 Carninci, P. and Hayashizaki, Y.
 High-efficiency full-length cDNA cloning
 Meth. Enzymol. 303, 19-44 (1999)
 MEDLINE
 PUBMED
 99279253
 10349636
 2
 REFERENCE
 3 Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Komno, H., Okazaki, Y., Muramatsu, M., and Hayashizaki, Y.
 Normalization and subcloning of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes
 Genome Res. 10 (10), 1617-1630 (2000)
 MEDLINE
 PUBMED
 20493974
 11042159
 4
 REFERENCE
 5
 TITLE
 RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer
 JOURNAL
 MEDLINE
 PUBMED
 20530913
 11076861
 AUTHORS
 Kawai, J., Shinagawa, A., Shibata, K., Yoshino, M., Itoh, M., Ishii, Y., Arikawa, T., Hara, A., Fukunishi, Y., Komno, H., Adachi, J., Fukuda, S., Aizawa, K., Izawa, M., Nishi, K., Kiyosawa, H., Kondo, S., Yamataka, I., Saito, T., Okazaki, Y., Gojobori, T., Bono, H., Kasukawa, T., Saito, R., Kadota, K., Matsuda, H., Ashburner, M., Batalov, S., Casavant, T., Fleischmann, W., Gaasterland, T., Gissi, C., Kind, B., Kochiwa, H., Kuehl, P., Lewis, S., Matsuo, Y., Nikaido, I., Pebole, G., Quackenbush, J., Schriml, L.M., Steubli, F., Suzuki, R., Tomita, M., Wagner, L., Washio, T., Sakai, K., Okido, T., Furuno, M., Aono, H., Baldarelli, R., Barsh, G., Blake, J., Boffelli, D., Bojunga, N., Carninci, P., de Bonaldo, M.F., Brownstein, M.J., But, C., Fletcher, C., Fujita, M., Gariboldi, M., Gustincich, S., Hill, D., Hofmann, M., Hume, D.A., Kamuya, M., Lee, N.H., Lyons, P., Marchionni, L., Mashima, J., Mazzarelli, V., Mombetis, P., Nordone, P., Ring, B., Ringwald, M., Rodriguez, I., Sakamoto, N., Sasaki, H.,

Sato, K., Schonbach, C., Seya, T., Shibata, Y., Storch, K.F., Suzuki, H.,
Toyo-oka, K., Wang, K.H., Weitz, C., Whitaker, C., Wilming, L.,
Wynshaw-Boris, A., Yoshida, K., Hasegawa, Y., Kawaji, H., Kohetsuki, S.
and Hayashizaki, Y.
Functional annotation of a full-length mouse cDNA collection
Nature 409 (6821), 685-690 (2001)

JOURNAL
MEDLINE
PUBMED
21085660
11217851

5
The FANTOM Consortium and the RIKEN Genome Exploration Research
Group Phase I & II Team.
Analysis of the mouse transcriptome based on functional annotation
of 60,770 full-length cDNAs
Nature 420, 563-573 (2002)

JOURNAL
AUTHORS
6 (bases 1 to 1896)

Adachi, J., Aizawa, K., Akahira, S., Akimura, T., Arai, A., Aono, H.,
Arikawa, T., Bono, H., Carninci, P., Fukuda, S., Fukunishi, Y.,
Furuno, M., Hanagaki, T., Hara, A., Hayatsu, N., Hiramoto, K.,
Hiraoka, T., Hori, F., Imotani, K., Ishii, Y., Itoh, M., Izawa, M.,
Kaukawa, T., Kato, H., Kawai, J., Kojima, Y., Konno, H., Kouda, M.,
Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Nishi, K.,
Nomura, K., Numazaki, R., Ohno, M., Okazaki, Y., Okido, T., Owa, C.,
Saito, H., Saito, R., Sakai, C., Sakai, K., Sano, H., Sasaki, D.,
Shibata, K., Shibata, Y., Shinagawa, A., Shiraki, T., Sogabe, Y.,
Suzuki, H., Tagami, M., Tagawa, A., Takahashi, F., Tanaka, T.,
Tejima, Y., Toya, T., Yamamura, T., Yasunishi, A., Yoshida, K.,
Yoshino, M., Muramatsu, M. and Hayashizaki, Y.
Direct Submission
Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of
Physical and Chemical Research (RIKEN), Laboratory for Genome
Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
Kanagawa 230-0045, Japan (E-mail: genome-res@gsc.riken.go.jp,
URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222,
Fax: 81-45-503-9216)
Please visit our web site (http://genome.gsc.riken.go.jp/) for
further details.

COMMENT
cDNA library was prepared and sequenced in Mouse Genome
Encyclopedia Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in RIKEN.
Division of Experimental Animal Research in Riken contributed to
prepare mouse tissues. First strand cDNA was primed with a primer
[5' GAGAGAGAGAGCGGCGGCAATCGAGTGTGTTTGTGTTTNN 3']. cDNA was
prepared by using trehalose thermo-activated reverse transcriptase
and subsequently enriched for full-length by cap-trapper. cDNA went
through one round of normalization to R0 = 5.0. Second strand cDNA
was prepared with the primer adapter of sequence [5'
GAGAGAGAGAGATCCAGAGCTCAATTAATTAATTAACCCCCCCC 3']. cDNA was
cleaved with XhoI and SstI. Cloning sites, 5' end: SstI; 3' end:
XhoI. Host: SOLR.

FEATURES
source
Location/Qualifiers
1..1896
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="FANTOM,DB:0610033E06"
/db_xref="MGI:1904280"
/db_xref="taxon:10090"
/clone="0610033E06"
/sex="male"
/tissue_type="Kidney"
/clone_lib="RIKEN full-length enriched mouse cDNA library"
/dev_stage="adult"
36..1629
/note="(UDP-glucuronosyltransferase 2 family, member 5
(MGI:MGI:98900)
putative"
/db_xref="MGI:98900"
BASE COUNT 571 a 361 c 386 g 578 t
ORIGIN
Query Match 52.4%; Score 864.2; DB 11; Length 1896;
Best Local Similarity 71.7%; Pred. No. 1,2e-17.;

Matches 1162; Conservative 0; Mismatches 453; Indels 6; Gaps 2;
QY 30 TCAGAGTGTCTAAGAAATGAGACTTACGCTCTTCCGATACAGCTGAGCTGTACTTGA 89
Db 31 TCAGAGTGTCTGGAAGATGATGCTCTCTGCTGCTGCGCAATAGTGTGCTGCTTCA 90
QY 90 GCTCTGGAAGTGTGGAAGGCTGTGTGCGCCACAGAAATTCAGCACTGATGATA 149
Db 91 GATCTGTGAATGTGGAAAGTGTGTGTGTGTGCGCGGATGAGATTCATGTGATGAACA 150
QY 150 TAAAGACATCTCGATGAACTTGTCCAGAGAGCTCATGAGTACTGTATGGCACTT 209
Db 151 TAAAGAAATATCTGAGAACTGTGTACAGAGGGCCATGAAGTACATCTGTCTGAGACTT 210
QY 210 CAGCTTCATTTCTTTCGATCCCAAGGCCATCTACTCTTAAATTTGAAGTTATCTTG 269
Db 211 CAGCTTCACTATGTTCTTGATCCCAAAATACCTGGCTTAAGTTGAAACCTTTTCTTA 270
QY 270 TATCTTAACTTAAACTGAGTTTGAGATATTTATCAAGCAGCTGTGTTAAGATGGGC-- 327
Db 271 CATCTGCATTAAGATGATCTGGAATTTTTCATACAGCTTTGAATGTGCACTT 330
QY 328 -AGAACTTCCAAAGACAACTTTGTGTCATATTTTTCACAAAGTACAAAGAAATCATGTGA 386
Db 331 ATGAGTTGTCAAGAGATACATGTTTGCATATTTCTCTTAATGCAAAATATGTTGATG 390
QY 387 CATTAATGACATACATTAAGAAAGTTCGTAGATATAGTTTCAATTAAGAAATCTTGA 446
Db 391 AATTATCTGTTTATTAATCTAAGCTTTGTGAAGATGCTTTCAACAGACAGCTCATGA 450
QY 447 AGAACTACAGAGAGTCAAGATTTGATGTTTCTTCAGAGATGCTTTCCCTTTGTTG 506
Db 451 CAAGCTACAGAGATCAAAATTTGATGCTCTTTGTGATGATCCTGTTGCTTTGTTGGTGG 510
QY 507 AGCTGTGCGCGAGTACTTAATAATCCCTTGTGTACAGCTTCGCTTCTCTGCT 566
Db 511 AGCTAATCGCTGAATTCCTCACATACCTTTCTGTACATGCTTGCCTTACATCAGATT 570
QY 567 ACGAATTTGAAAGCATAGTGAAGAGATTTCTGTTCCCTCTCTTATGTCCTGTTGTTA 626
Db 571 ACGAATTTGAAATTCAGAGTGAAGATTTCTACTACCTCCCTTATGTCCTGTTATTC 630
QY 627 TGTGAGAACTAAGAGCAATATGCTTTCAATAGAGAGGTAAATAATGATCATGTGC 686
Db 631 TGTGAGATTAAGTGAAGCAATATGATCATTAAGAGAGATTAATAATATGATGATGC 690
QY 687 TTTATTTGAATTTGTGTTCCAAATATTTGATGATGAAGAGTGGATCACTTACAGTG 746
Db 691 TTTATTTGAATTTGTGTTCCAGATGCTTAATGATTAAGAAATGGATTTCAATTTACACTG 750
QY 747 AAGTTCAGAGAACCCACTACATGTTATCTGAGCAATGCGAAAGTGCATATAGCTTTA 806
Db 751 AATATTTGGAAAGGCCACACACCTTACTAGAACAAAGGGCCAAAGCAAGAAATGTGGCTCA 810
QY 807 TTGCAAACTATGGGATTTTCAATTTCTGACCCACTTTACCAAAATGTTGAGTTGCTTG 866
Db 811 TTGATCAACTGGGATTTTGAAGTTTCTCAACCCGACCTTACCAAAATGTTGATGATGTTG 870
QY 867 GAGGACTCACTGCAAACTGTCCAAACCCCTACCGAAGAAATGGAAGAGTTGTCCAGA 926
Db 871 GAGGACTCACTGCAAACTGTCTTAACTTGTCTTAAGATATGGAAGAAATATATGTCCAGA 930
QY 927 GCTTGTGAAGAAATGTGTGTGTGTGTTTCTCTGCGGTGAGATGTCATGAACGTCAG 986
Db 931 GCTTGTGAAGCAATGTGTGTGTGTGTTTCTCTGCGGTCAATGTGTGAATCAATCAAG 990
QY 987 AAGAAAGGCAATGTATGTGATCATGAGCTTGTGCAAGATCCCAAAAGATTTCTGTGGA 1046
Db 991 AAGAAAGGCAATGTATGTGATCATGAGCTTGTGCAAGATTTCCAAAGAGTTCTTTGGA 1050
QY 1047 GATTGTAGGGAATTAACAGATATCTTGAAGTCAATCTGCGCTGTACAGTGTATAC 1106
Db 1051 AATTGTATGCAAAACCCAGCAACTTTAGAGACAAATATACAGAGTGTACAGTGTCTCC 1110


```

FEATURES
  source
      Location/Qualifiers
          1..1783
              /organism="Mus musculus"
              /mol_type="mRNA"
              /strain="C57BL/6J"
              /db_xref="PANTOM_DB:C730037110"
              /db_xref="taxon:10090"
              /clone="C730037110"
              /sex="male"
              /cell_type="tumor"
              /issue_type="liver"
              /clone_id="RIKEN full-length enriched mouse cDNA library"
              /dev_stage="adult"
          34..1638
              /note="unnamed protein product; putative
              similar to UDP GLUCURONOSYLTRANSFERASE UGT2A3 [Cavia
              porcellus] (SPTRIQ9R10, evidence: FASTA, 71.3%id,
              99.8%length, match=1579)"
              /codon_start=1
              /protein_id="BAC34191.1"
              /db_xref="GI:26341058"
          /translation="MVSEKCYAIFLLQLCAGGCFSCXYLWPCDMSHMLNLTILE
          ELGARGHEVVLKPSIIIDSKRIPLHFNITPLVEIETAKNLNLIANLAVNIN
          LSLMEAKTIQDFLOVTDIESI.CRSVLNOKFMDLRDQYDVVVIDEVPQGEIV
          AEVLQIFVYTLRFSSMGYMEKHGOLPIPLSYVPMSELDMTEFERKMMFSL
          LFEVWLQOYDPAFMDOYSETLGRPTTECTGKADIMLRTYMDVEFPPLYPNEF
          VGGIHCBAKPLKEMEEFVOSGEHVVVPSLSGMKYNLTREKANIASVLAQIPK
          VUNRSKRPATLGSNTRLNWIRIQNLLHPTKAITTGGTNGIYEAHYGVPMVG
          VPMIGDQPHNIAHMEAKGAAIKVSISTMSTDLISAVRAVINESYENMRISRIH
          DQPKPLRAVFIWFYFMRHKGAHLRVAHDLISWFOYHSLDIVGLLVCVLLFTPI
          TKFLFVCKLYMKESKKMGNRKKKN"
          1756..1761
              /note="putative"
          polyA_signal
              1783
              /note="putative"
          polyA_site
              1783
              /note="putative"
BASE COUNT      510 a      365 c      416 g      492 t
ORIGIN
Query Match      47.6%; Score 785.8; DB 11; Length 1783;
Best Local Similarity 69.9%; Pred. No. 4e-155;
Matches 1119; Conservative 0; Mismatches 472; Indels 10; Gaps 4;
QY      28 CATCAGATGCTCTAATGAAATGACCTCAGCTCTCTCTGATACAGCTGAGCTTACTT 87
Db      30 CAATATGCTCTCTAATAATGCTTGCGGCACTTTTCTGCTGACGCTTGCTGGCC-- 87
QY      88 TAGCTCGGAGTGTGGAAGGTGCTGTGTGCGCCACAGAAATTCAGCCACTGATGAA 147
Db      88 -GGCTGTGATTCGACGAAGGTCTCTGTGTGCGCCGTGTGATGAGCCACTGGCTGAA 146
QY      148 TATTAAGCAATCTGTGATGAACTTGTTCAGAGGCTCATGAGTGAAGTGTATGGCATC 207
Db      147 TCTAAGACATATCTTGAAGGCTTGGAGCAAGGCGCACAGAGTAAACAGTCCGAAATA 206
QY      208 TTCAGCTTCATTTCTTTCATCCCAACAGCCCACTACTCTTAATTTGAAGTTATCC 267
Db      207 CCCCAATATCAT--CATATGACAGATGAACGATTCCTCACTGCACTTTGGAATATCC 263
QY      268 TGTATCTTAACTAATAAATGATTTGAGATATATCAAGCAAGCTGTTAAGAGA--TG 324
Db      264 TTGCTGTATGAATTCAGACAGCTGAGAAATCGTTAATGATGCAAAATCTAGCTGT 323
QY      325 GCGGAACCTTCCAAAGACACATTTTGTGCTATATTTTCAAGTACAGAAATCATGTG 384
Db      324 GAATGTCATTCCAACCTGTCTACGTGGGAGACAGCAAAACATTAACAGCTTCTTCT 383
QY      385 GACATTAATGACATCTAGAAAGTCTGTGAGATATAGTTCAATATGAAGAACTTAT 444
Db      384 TCAAGTAACTGAGATTTTGAAGTATTTTGAAGAGTATTTGACACAGAAATTCAT 443
QY      445 GAAGAAATCAAGAGTCAAGATTTGATTTGTTCTTGCAGATCTGTTTCCCTTTGG 504
Db      444 GGAACAACTACGGGATCACAAATATGATGTAGTGTATAGACCTGTCTGTTCCCTGTG 503

```

```

QY      505 TGAGCTGTGCGCAGATTACTTAATAACCTTTGTCTACAGCCTCGCTTCTCTCGG 564
Db      504 AGAGTTGTGCGCAAGAGTGTCTTACATCTCTTCTGATACACATGAGTTACAGATGG 563
QY      565 CTACGCAATTGAAAGCATAGTGAAGACTTGTTCCTCTCTCTATGTGCTGTGT 624
Db      564 CTACTACATGAGAAACAGTGTGCGCCTTCCAAATTCACCTCTGATGACCGGTTGT 623
QY      625 TATGTCAAACTAAGTACCAATGACTTTTCATAGAGAGGTAAATAATGATCTATGT 684
Db      624 CATGAGTACACTACAGCAAAATGTGACCTTACAGAGAGGTGAAAATATGATGTTTC 683
QY      685 GCTTATTTTGAATTTTGTTCCTCAAAATTTTGAATGACATGAAGAAGGGATCTACAG 744
Db      684 ACTGTTGTTGAGTACTGCGCTCCAGCATATGACTTTGCACTTGGGATCAGTTTACAG 743
QY      745 TGAAGTTCTAGAAAGCCCACTACGTTATCTGAGACATGACAAAGCTGACATAGGCT 804
Db      744 TGAACCTCTAGAAAGGCCCAACGTTCTGTAAAGCTGTGGGAAAGCTGACATTTGGCT 803
QY      805 TATTCGAATCTACTGGGATTTTCAATTTCTCACCCACTTTACCAATGTTGAGTTGCT 864
Db      804 AATCCGAACATATTTGGGATGTGAGTTTCTCTGCTCATATTTACCAATTTTGAGTTGT 863
QY      865 TGGAGGACTCCACGCAAACTGCGCAACCCCTACGGAAGAAATGGAAGATTTGTCCA 924
Db      864 GGAAGACCTCACCTGCMAACCTGCAAGCTTTTACTTAAGAAATGGAATTTGTTCA 923
QY      925 GAGCTCTGAGAAATATGTTGTGTGTGTGTTTCTCTGGGGTCGATGTCATGACAGTC 984
Db      924 GAGCTCTGAGAAACATGTGTATGATATTTTCACTGGGGTCAATGATCAAAAACCTGAC 983
QY      985 AGAAGAAAGGCCCAATGTATTTGATCAGCCCTTGGCCAAATCCCAAAAGGTTCTGTG 1044
Db      984 AGAAGAAAGGCCCAACTCATTCCTCTGTCTCTGCGCCAAATTTCCAGAAAGGTTGTG 1043
QY      1045 GAAATTTGATGGGAATTAACCAATGCTTTAGAGCTCAATACCTGGGTGACAAAGTGGAT 1104
Db      1044 GAGATTACTAGGCAAGAGAGCCAGCCACATTAAGATCCAAATACCTGGGTTTAAATGGAT 1103
QY      1105 ACCCGAATGATCTTCTTGTGATCCCAAAACCAAAAGCTTTTATCATCTATGGTGAAT 1164
Db      1104 TCCCGAATGATCTTCTTGTGATCTTAAACCAAAAGCTTTTATCATCATATGTTGAAC 1163
QY      1165 GAATGGATCTATGAAGCTATTTTCAATGAGGGTCTCTATGTGAGAGTTCCTATTTGG 1224
Db      1164 AAACGGGATTTATGAAGCCATTTACATGAGGGTCTCTATGTGAGGCTTCCCATGTTAGG 1223
QY      1225 TGATCAGCTTGAATACATAGCTCAATGAGGCGAAAGGACAGCTGTAAATTAACCTT 1284
Db      1224 GATCAGCTTCAACCAATGCTGCAATGAGGCGCAAGGGACAGCCCTGAAAGTCAGCAT 1283
QY      1285 CAAACATATGACAAGGAGATTTTACTGAGGGCTTTGAGAAAGCTTATCCGATTCCTC 1344
Db      1284 CAGTACATATGACAGACAGATTTTACTCAGTGTCTGTGAGGCGATATTAATGACCTTC 1343
QY      1345 TTATTAAGAGATGCTATGAGATTTATCAGAAATTCACCATGATCAACTGTAAAGCCCT 1404
Db      1344 TTATTAAGAGATGCTATGAGATTTATCAGAAATTCACCATGATCAGCACTGAAGCCCT 1403
QY      1405 AGATCGACAGCTCTTCTGATGAGTGTGTCAGGCGCAAAAGGAGCCAGACCTGGG 1464
Db      1404 GAGCCGACAGCTCTTCTGATGAGTGTGTCAGGCGCAAAAGGAGCCAGACCTTTGG 1463
QY      1465 ATCAGTGCCTCATGACCTCACTGTTCCAGCACTACTATAGATGATGATGGTTCCT 1524
Db      1464 TGTGAGAGCCCAAGACCTCACTGTTTCACTGATGACACTCCCTAGATGATGATGGTTCCT 1523
QY      1525 GCTGACCTGTGTGCAATGCTATATTTCTTGTTCACAAATGTTTTTATTTTCTGTCA 1584
Db      1524 ATTGTGTGTGTGTACTCTGACATTCATCACTCAATTAATTTTGTGTGTGTGTCA 1583

```

| | | | |
|----|------|--|------|
| Oy | 1585 | AAATTTTATAACTAGAAAATGATGAAGAAGGAAATGA | 1625 |
| Dd | 1584 | AAAACTTATTATGA-AAGAAAATAAGAAAATCGGGAACGA | 1623 |

| RESULT 7 | AKO08601 | 2099 bp | mRNA | linear | HTC 05-DEC-2002 |
|-----------|---|-------------|---|--------|-----------------|
| LOCUS | AKO08601 | | Mus musculus adult male small intestine cDNA, RIKEN full-length enriched library, clone:2010331U07 product:similar to UDP GLUCURONOSYLTRANSFERASE UGT2A3 [cavia porcellus], full insert sequence. | | |
| ACCESSION | AKO08601 | | | | |
| VERSION | AKO08601.1 | GI:12842684 | | | |
| KEYWORDS | HTC; CAP trapper. | | | | |
| SOURCE | Mus musculus (house mouse) | | | | |
| ORGANISM | Mus musculus | | | | |
| REFERENCE | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. | | | | |
| AUTHORS | Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Kono,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y. | | | | |
| TITLE | High-efficiency full-length cDNA cloning | | | | |
| JOURNAL | Meth. Enzymol. 303, 19-44 (1999) | | | | |
| MEDLINE | 99279253 | | | | |
| PUBMED | 10349636 | | | | |
| REFERENCE | 2 | | | | |
| AUTHORS | Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Kono,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y. | | | | |
| TITLE | Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes | | | | |
| JOURNAL | Genome Res. 10 (10), 1617-1630 (2000) | | | | |
| MEDLINE | 20499374 | | | | |
| PUBMED | 11042159 | | | | |
| REFERENCE | 3 | | | | |
| AUTHORS | Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P., Komno,H., Akiyama,J., Nishi,K., Kitsuana,T., Teshito,H., Itoh,M., Suni,N., Ishii,Y., Nakamura,S., Hazama,M., Nishie,T., Harada,A., Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kasaiwagi,K., Fujiwara,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watabiki,M., Oyake,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsunura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y. | | | | |
| TITLE | RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multipillarary sequencer | | | | |
| JOURNAL | Genome Res. 10 (11), 1757-1771 (2000) | | | | |
| MEDLINE | 20530913 | | | | |
| PUBMED | 11076861 | | | | |
| REFERENCE | 4 | | | | |
| AUTHORS | Kawai,J., Shinagawa,A., Shibata,K., Yoshino,M., Itoh,M., Ishii,Y., Arakawa,T., Hara,A., Fukunishi,Y., Konno,H., Adachi,J., Fukuda,S., Aizawa,K., Izawa,M., Nishi,K., Kiyosawa,H., Kondo,S., Yamanaka,I., Saito,T., Okazaki,Y., Gojobori,T., Bono,H., Kasukawa,H., Saito,R., Kadota,K., Matsuda,H., Ashburner,M., Batalov,S., Casavant,T., Fleischmann,M., Gaasterland,T., Glass,C., King,B., Kochiya,H., Kuell,P., Lewis,S., Matsuo,Y., Nikaido,I., Pesole,G., Quekembush,J., Schriml,L.M., Staudt,F., Suzuki,R., Tomita,M., Wagner,L., Washio,T., Sakai,K., Okido,T., Furuno,M., Aono,H., Balarelli,P., Barsh,G., Blake,J., Botfield,D., Bojunga,N., Carinici,P., de Bonaldo,M.F., Brownstein,M.J., Bul,C., Hofmann,C., Fujita,M., Gariboldi,M., Gusticich,S., Hill,D., Pfeiffer,M., Hume,D.A., Kamiya,M., Lee,N.H., Lyons.P., Marchionni,L., Mashima,U., Mazzerelli,J., Mombaerts,P., Nordone,P., Ring,B., Ringwald,M., Rodriguez,I., Sakamoto,N., Sasaki,H., Sato,K., Schonbach,C., Seya,T., Shibata,Y., Storch,K.F., Suzuki,H., Toyoh-oke,K., Wang,K.H., Weitz,C., Whitaker,C., Wilming,L., Wynshaw-Boris,A., Yoshida,K., Hasegawa,Y., Kawaji,H., Kohsaki,S. and Hayashizaki,Y. | | | | |
| TITLE | Functional annotation of a full-length mouse cDNA collection | | | | |
| JOURNAL | Nature 409 (6621), 685-690 (2001) | | | | |
| MEDLINE | 21085660 | | | | |
| PUBMED | 11217851 | | | | |
| REFERENCE | 5 | | | | |
| AUTHORS | The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II team. | | | | |

TITLE
JOURNAL REFERENCE
AUTHORS

Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs
Nature 420, 563-573 (2002)
(bases 1 to 2039)

Adachi,J., Aizawa,K., Akahita,S., Akimura,T., Arai,A., Aono,H., Arakawa,T., Bono,H., Carninci,P., Fukuda,S., Fukushima,Y., Furuno,M., Hanagaki,T., Hara,A., Hayatsu,N., Hizemoto,K., Hiraka,T., Hoti,F., Imocani,C., Iihii,Y., Itoh,M., Izawa,M., Kasukawa,T., Kato,H., Kawai,J., Kohima,Y., Konno,H., Kouda,M., Koya,S., Kuithera,C., Matsuyama,T., Miyazaki,A., Nishi,K., Nomura,K., Numezaki,R., Ohno,M., Okazaki,Y., Okido,T., Owa,C., Saito,H., Saito,R., Sakai,C., Sekai,K., Sano,H., Saeki,D., Shibata,K., Shibata,Y., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F., Tanaka,T., Tejima,Y., Toya,T., Yamamura,T., Yasunishi,A., Yoshida,K., Yoshino,M., Muramatsu,M. and Hayashizaki,Y.

TITLE
JOURNAL

Direct Submission
Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp,
URL:http://genome.gsc.riken.go.jp/, Tel.:81-45-503-9222,
Fax:81-45-503-9216)

COMMENT

Please visit our web site (<http://genome.gsc.riken.go.jp/>) for further details.
cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of genome exploration research group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. First strand cDNA was primed with a primer [5'-GAGGAGGAGAAGCATCAAGCGCTCTTTTCTTTTCTTNN 3'], cDNA was prepared by using trehalase thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of normalization to Rot = 5.0 and subtraction to Rot = 20.0. Second strand cDNA was prepared with the primer adapted of sequence[5'
GAGGAGGACATCTTCGAGTTAATTAAATTAATCCCCCCCCCCC 3']. cDNA was cleaved with XhoI and SacI. Cloning sites, 5' end: XhoI, 3' end: SacI.
Host: SOLR.

Location/Qualifiers

1..2039
/organism="Mus musculus"
/mol_type="mRNA"
/set_ain="CS7BL/6J"
/db_xref="FANTOM,DB:201032JU07"
/db_xref="MGI:1904826"
/db_xref="taxon:100930"
/clone="201032JU07"
/sex="male"
/tissue type="Small intestine"
/clone_id="RIKEN full-length enriched mouse cDNA library"
/dev stage="adult"
15..1619
/note="unnamed protein product; putative similar to UDP GLUCURONOSYLTRANSFERASE UGT2A3 (Cavia porcellus) (SPTRLQ9RLD,evidence:FASTV, 71.3%ID, 99.8%length,match=1579)"
/codon_start=1
/protein_id="BAB25770.1"
/db_xref="GI:12842885"
/db_xref="MGI:1919144"
/translation="MVSEKCYAAPFLIQLCMAGCGFGSKYLVMPCKDSHWMLTKLITLLELGARGHEVTLPKPSLIIDSKRIPLHFEIVPLILEETREPLNLANIANLVIPVCELSLEAEAKTLQDFLOQTGDPEISICRYLVNQMKDKRDQOVVVIVDPVPCGLAEVQILPVEYLFIRSMGVMEKHGCGLPIPSVPVPMSELDMTPFERVKMMFSLIEFWMLQVDYPALRMDOYSERTIGRTPEFKVGADILWLRITWDVEPRPYLLPFEEFEGVGHCHCRAPKLPREMEEPFNSGEHVAVSPKSMYTNLEEKANIALSYLAOIPOIVLMVSGKPAILGSNTRELFWMTITQNDLGHPRKTATTHGCTNGCIAYAIYGPVMNVPMGGDDPHIAHWEARGAALKVISITWTSTDLISARAVINBSPYKEMARSLSIHHDOPKPLDRAVFVKEIFVMRHKGAKRLVAADLSWFOYHSLDVIGFLCVLTFTFIITKFCFLPCKLYKMSKKMGKRKKKN"

ORIGIN

47.38; Score 781; DB 11; Length 2099;

| | | | | |
|---------------------------|-------|---------------------|------------|--------------|
| Query Match | 47.3% | Score 781; | DB 11; | Length 2099; |
| Best Local Similarity | 69.7% | Pred. No. 4.2e-154; | | |
| Matches 116; Conservative | 0; | Mismatches 475; | Indels 10; | Gaps 4; |

| | | | | |
|----|-----|---------------------------|--|------|
| QY | 28 | TATCAGAGTGTCTATGAAATGGA | CTTCAAGCTCTCTCGTAAATGACGTGAGCTGTACTT | 87 |
| Db | 11 | CAATATGCTCTTGAAAATATGTTGG | CGCAATTTTTTCTGCTGCAGCTTTGCTGGCC-- | 68 |
| QY | 88 | TAGCTCTGGGAGTTGTGGAAAGTG | CTGCTGCTGGCCCAAGAAATTCAGCCACTGGATGA | 147 |
| Db | 69 | -GGCTGTGGATTCTGCAGCAGAGT | CTCTGTGTGGCCCTGTATATGACCACTGGCTAA | 127 |
| QY | 148 | TATTAAGACATCTCTGATGAACTT | GTCCAGAGAGCTATGAGTGACTGTATTTGGCATC | 207 |
| Db | 128 | TCCTAAAGCTATTTCTTTGAGAGCT | TGGACAAAGGGGACAGAGGTAAACAGTCTGTAAATA | 187 |
| QY | 208 | TTGAGCTTCACATTTCTTTGCAT | TCCCAAGCCCATCTACTCTTAAATTTGAAAGTTATCC | 267 |
| Db | 188 | CCCCAGTATCTATCTGAGTATCT | GCATGATTAAGTAAACGATTTCACTGCACITTTGAGAAATATTC | 244 |
| QY | 268 | TGTATCTTTAACTAAACCTGAT | TGAGATATATCAAGCAGCTGTGTTAAGAGA--TG | 324 |
| Db | 245 | TTTCTGTATGAATTCAGACAGCT | GAGATTCGTTAAATGAGATGCAAAATCTAGCTGT | 304 |
| QY | 325 | GGCAGAACTTCCAAAAGACACAT | TTTGGTCAATTTTTCACAAGTACAGAATATAGTG | 384 |
| Db | 305 | GAATGTCAATTCCAAACCTGTCA | CTGTGGAAAGCAGCAAAAACATTTACAAAGCTTCTTCT | 364 |
| QY | 385 | GACATTTAATGACATCTTAGAAAG | TTGTGTATGAGATTTCAATTAAGAAACTAT | 444 |
| Db | 365 | TCAAGTAACTGGAATTTTGAAGA | ATTTGTATGAGATATTTGTAACAACGAAATTCAT | 424 |
| QY | 445 | GAAGAAACTACAGAGTCAGATTT | GATGTGTTCTTGACAGATGCTGTTTTCCCTTTGG | 504 |
| Db | 425 | GGACAACTACGGGATGCACAAT | ATATGATGTGTTATTAAGCCTGTGCTTCCTGTGG | 484 |
| QY | 505 | TGAGCTGTGGCCGAGTTACTTTAA | ATATACCTTTGTCTACAGCTTCGCTTCTCTCCGG | 564 |
| Db | 485 | AGAGTTGGTGCAGAAAGTCTTCA | GAATCCCTTTCGTATACACACTGAGGTTTCAGCATGG | 544 |
| QY | 565 | CTAGGCAATTAAGACATAGTGA | GAGACTCTGTGCCCTCTCTCTTAATGAGCTGTGT | 624 |
| Db | 545 | CTACTACATGAGAAACACTGTGG | CCAGCTTCCAAATTCACCTCTGTATATGACGGTTGT | 604 |
| QY | 625 | TATGTCAAACTAAAGTGAACAA | TATGACTTTTCAATAGAGGGGTAAAAAAATATGATCTATGT | 684 |
| Db | 605 | CATAGTAGAGCTGACAGACAT | ATATATGACCTTTCACAGAGAGGGGTGAAAAATATGATGTTTC | 664 |
| QY | 685 | GCCTTATTTTGAATTTTGGTCC | AAATATTTTGAACATGAAAGAGGGGATTCAGTTCTAAG | 744 |
| Db | 665 | ACTGTGTGTGAGTACTGTGCTC | ACAGCAATATATGACTTTGCAATTCGGGATCAGTTTAAAG | 724 |
| QY | 745 | TGAAGTCTAGGAAGACCCCTAG | GTATCTGAGACATGACAAATGACAAAGCTGACATATGCT | 804 |
| Db | 725 | TGAAACCTTAGGAAGGCCCA | CAACGTTCTGTAAAGCTGTGGGGAGAGCTGACATTTGGCT | 784 |
| QY | 805 | TATTCGAAACTACTGGGATTTT | CAATTTCTTCACCCCACTCTTACCAATGTTGAGTTGCT | 864 |
| Db | 785 | AATTCGAAACATTTGGGATGT | TGAGTTTCTTCGTCCATATATTAACCAATTTTGGAGTTGT | 844 |
| QY | 865 | TGAGAGCTCCACTGCAAACTG | CAAAACCCCTACCGAAGGAATGGAAGAGTTGTGCA | 924 |
| Db | 845 | GGGAGAGTGCACGTCAAACTG | CCAAAGCTTTTACTTAAGGAATATGGAAGAAATTTGTCA | 904 |
| QY | 925 | GAGCTCTGAGAAAATGCTGT | TGTGTGTTTTTCTCTGGGATCGATGTCAAGTAAACAGCTC | 984 |
| Db | 905 | GAGCTCTGAGAAACATGGTGT | ATAGTATTTTCACTGGGGTCAATATGCTCAAAAACCTGAC | 964 |
| QY | 985 | AGAGAAAGGGCCATATTTG | ATTCATGCAAGCCCTTGCAAGATCCCAAAAAGGTTCTGTG | 1044 |
| Db | 965 | AGAGAGAAACCACTCACTG | CTCTGTCTTGTGCCAGATATCCCAAGAAAGGTTTGTG | 1024 |

| | | | |
|----|------|---|------|
| QY | 1045 | GAGATTGATGGGAAATTAACAGATCTTTAGAGCTCAAACTCGGCTGACAACTGAT | 1104 |
| Db | 1025 | GAGATTCTCAGGCCAGAGAGCCAGCATATTAGATTCAAATCTGGCTTTTAAATTGGAT | 1084 |
| QY | 1105 | ACCCAGAAATGATCTTTTGGTCATCCCAAAACCAAGCTTTTATCACTCACTGGTGAAT | 1164 |
| Db | 1085 | TCCCAAGATATGATCTTCTTGAGACATCTTAAACCAAGCTTTTATCACTCACTGGTGAAC | 1144 |
| QY | 1165 | GAATGGGATCTATGAGACTATTTACATGGGGTCCCTATGGTGGAGTTCCCATATTTGG | 1224 |
| Db | 1145 | AAAGGGATTTATGAAACCATTTACATGGGGTCCCTATGGTGGCGTTCCCATGTTAGG | 1204 |
| QY | 1225 | TGATCAGCTTGATATACATAGCTCACTAAGAAAGGCCAAAGAGCAGCTGTAGAAATTAACCT | 1284 |
| Db | 1205 | GGATCAGCTCTCAACATCGCTCACTAGAGGCCAAGAGAGCAGCCCTGAAAGTCAGCAT | 1264 |
| QY | 1285 | CAAAATCTATGACAAGCCAAAGATTATCTGAGGGCTTTGAGACAAGTCAATTAACCGATTCCCTC | 1344 |
| Db | 1265 | CAGTACAAATACGAGCACAATTTTACTCAGTGCCTGTAGGGCAGTATTTAATGAGCCTTC | 1324 |
| QY | 1345 | TTTATTAAGAGATGCTATGAGATTATCAAGAAATTCACCATGATCACTCTGTAAGCCCT | 1404 |
| Db | 1325 | TTTATTAAGAAATGCCATGCGGTTATCAAAATTCACCATGATCAGCAGTGAAGCCCT | 1384 |
| QY | 1405 | AGATCGAGCAGCTCTTCGAGATCGAGTTTGTCATGCGCACAAGAGAGCCAAAGCAGCCTGGC | 1464 |
| Db | 1385 | GGACAGCAGCAGCTCTTCGAGATGAGTTTGTCATGCGCACAAGAGAGCCAAAGCAGCCTTGC | 1444 |
| QY | 1465 | ATCAGCTGCCCATGACCTCACTGGTTCACAGCACTACTCTATAGATGATGATTGGGTTCT | 1524 |
| Db | 1445 | TGAGCAGGCCCATGACCTCACTGGTTCACAGCACTACTCTATAGATGATGATTGGGTTCT | 1504 |
| QY | 1525 | GCTGACCTGTGTGGCAACTGCTATATTCTTGTTCAAAAATGTTTATTTTCTTGCTCA | 1584 |
| Db | 1505 | ATTGTTGTGTGCTTACTGTCATTCATCATCAATTAATTTTGTGTTGTGTGTCA | 1564 |
| QY | 1585 | AAAAATTAATTAACCTGAAAGATAGAAAGAGGAGGATTAAC | 1625 |
| Db | 1565 | AAAACTTTATATGA-AGAAAGTAAAGAAATGGGAAACA | 1604 |

| RESULT | 8 | | | | |
|------------|--|-------------|-------|--------|-----------------|
| EC048920 | | | | | |
| LOCUS | | | | | |
| DEFINITION | BC048920 | 2575 bp | mRNA | linear | HTC 27-MAR-2003 |
| ACCESSION | Mus musculus, clone IMAGE:6311381, | | mRNA. | | |
| VERSION | BC048920 | | | | |
| KEYWORDS | BC048920.1 | GI:29294723 | | | |
| SOURCE | HTC. | | | | |
| ORGANISM | Mus musculus (house mouse) | | | | |
| REFERENCE | Mus musculus | | | | |
| AUTHORS | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. | | | | |
| TITLE | 1 (bases 1 to 2575) | | | | |
| JOURNAL | Strausberg,R. | | | | |
| REMARK | Direct Submission | | | | |
| COMMENT | Submitted (17-MAR-2003) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA | | | | |
| | NIH-MGC Project URL: http://mgc.nci.nih.gov | | | | |
| | Contact: MGC help desk | | | | |

through the I.M.A.G.E. Consortium/LML at: <http://image.lml.gov>
 Series: IRAX Plate: 105 Row: f Column: 14
 This clone was selected for full length sequencing because it
 passed the following selection criteria: Hexamer frequency ORF
 analysis

This clone has the following problem: frame shifted.

FEATURES

source

Location/Qualifiers
 1..2575
 /organism="Mus musculus"
 /mol_type="mRNA"
 /db_xref="taxon:10090"
 /clone="IMAGE:6311381"
 /issue="type="Olfactory epithelium, neonatal mouse,
 C57Bl/6"
 /clone_id="N1H MGC_129"
 /lab_host="DH10B"
 /note="vector: PCMV-SPORT6"

BASE COUNT 799 a 492 c 519 g 765 t
 ORIGIN

Query Match 46.6%; Score 769; DB 11; Length 2575;
 Best Local Similarity 69.8%; Pred. No. 1.4e-151;
 Matches 1085; Conservative 0; Mismatches 460; Indels 10; Gaps 3;

```

102 GTGGAAGTCTGTGTGGCCACAGATTTCAGCCAGTGTGATATATAAGACATCC 161
113 GCGGGAATGTGTATTTGGCTTACAGATGGCAGCATTTGGTAAATTAAGATTCTTC 172
162 TGGATGAATGTGTCCAGAGAGTCAATGAGTGTATGATGATGATGATGATGATGAT 221
173 TTGAGAGTGTGTGTCAAGAAATCAAGTGTGATGATGATGATGATGATGATGATGAT 232
222 GTTGTGATCCCAAGAGCCATCTACTTAAATTTGAAGTTATCTGATCTTAACTA 281
233 TTATTAATCTCCAGCTTGTATGCTTTATTAATTTGAAGATTTCCGTTTGTACAA 292
282 AAAGTGTGTGAGATATTTATCAAGCAGCTGTGTTAAGATGGGACAG-----ACTTC 335
293 AAAGCAATGATGATTAATTAATTTGAGCATATGATGACATGTGGCTGGACCAAGCCAA 352
336 CAAAAGACATTTTGTGTATTTTTCACAAATCAAGAAATCAATGTGACATTTAATG 395
353 GCGCTCTCAATGTGACCTTTCTCAAAAGAACTAGAAATCTCTTGTCTTCTTCA 412
396 ACATCTTATGAAAGTCTGTAGATATGATTTCAATTAAGAACTTATGAAAGAACTAC 455
413 CAACCAATTAAGCAATGTGTATGATGATTAACCAATCTCACTGTAATGAAAGAACTGC 472
456 AGAGTCAAGATTTGATGTTTGTGTGAGATGCTGTTTCCCTTTGGTGAAGTGTGG 515
473 AAAAAGGTGCTTCAATGTGTATGATGATGATGATGATGATGATGATGATGATGAT 532
516 CCGAGTTACTTAATAATACCTTTGTCTACAGCTTCCGCTTCTCTCTGCTACGAATGG 575
533 CTCTGAAATTTGGGAGATTCCTTTGTGTATACACTACGCTTCTCTCAGCCTTTAAGTGG 592
576 AAAACATGTGTGAGAGATCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 635
593 AAGAGCATTTGTGGAAATTTCTGCCCCCATCTCACTACCTACCTGAGCCCTGTACAGC 652
636 TAAAGTCAAGATTTCTCATAGAGAGGTAAATAATGATGATGATGATGATGATGATGAT 695
653 TCACTGACCAATGTCTTTTGTGTGAAAGGTTTAAGAAATCAATTTCTTATTTCTGTGCAAG 712
696 AATTTGTGTTCAAAATATTTGACATGAAGAGTGGATGATGATGATGATGATGATGATGAT 755
713 ACTATATATTTAAAGCT---ACTGGGAGATGATGATGATGATGATGATGATGATGATGAT 769
756 GAAAGCCCACTACGTTATCTGAGACAT--GCAAAAGCTGACATATGCTTATTTGGAAC 814
770 GAAAGCCCACTACGTTATCTGAGACAT--GCAAAAGCTGACATATGCTTATTTGGAAC 829
815 TACTGGATTTTCAATTTCTCTCAACCACTTACCAATGTTGAGTTGCTTGGAGAGATCTC 874

```

```

830 TATGGATTTTGAATTTCTCTGCTCAATTTACCAATTTGATGTTGGGAGACTG 889
875 CACTGCAACCTGCAACCCCTTACGAGAAATGGAAGTGTGCTCAGAGCTTGA 934
890 CACTGCAACCTGCAACCCCTTACGAGAAATGGAAGTGTGCTCAGAGCTTGA 949
935 GAAATGATGTTGTGTGTTTCTCTGAGGTCATGTGATGATGATGATGATGATGATGAT 994
950 GAACATGATGTTGTGTGTTTCTCTGAGGTCATGTGATGATGATGATGATGATGATGAT 1009
995 GCCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1054
1010 GCCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1069
1055 GGGATTAACAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1114
1070 GGAAGATTAACAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1129
1115 GATCTCTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1174
1130 GATCTCTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1189
1175 TATGAGCTATTTACATGAGGTCCTATGTGTGAGGATGATGATGATGATGATGATGATGAT 1234
1190 TATGAGCTATTTACATGAGGTCCTATGTGTGAGGATGATGATGATGATGATGATGATGAT 1249
1235 GATTAACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1294
1250 GACAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1309
1295 ACAAGCAAGATTTACTGAGGTCCTTATGAGACATGATGATGATGATGATGATGATGATGAT 1354
1310 ACAAGCAAGATTTACTGAGGTCCTTATGAGACATGATGATGATGATGATGATGATGATGAT 1369
1355 AATGCTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1414
1370 AATGCTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1429
1415 GTCCTTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1474
1430 GTCCTTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1489
1475 CATGACCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1534
1490 CATGACCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1549
1535 GTGGCACTGCTATATTTCTGTTCAAAATGTTTATTTTCTGTTCAAAATTTAT 1594
1550 GTGGCACTGCTATATTTCTGTTCAAAATGTTTATTTTCTGTTCAAAATTTAT 1609
1595 AAAACTGAAATGAAAGAAAGAGGATGATGATGATGATGATGATGATGATGATGATGAT 1649
1610 AAGACAGGAAAGAAAGAAAGAGGATGATGATGATGATGATGATGATGATGATGATGAT 1664

```

RESULT 9
 BX444042 983 bp mRNA linear EST 15-MAY-2003
 LOCUS BX444042 Homo sapiens FETAL LIVER Homo sapiens cDNA clone
 DEFINITION CS0DM005Y05 5-PRIME, mRNA sequence.
 ACCESSION BX444042
 VERSION BX444042.1 GI:30778198
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 983)
 AUTHORS Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
 TITLE Full-length cDNA libraries and normalization
 JOURNAL Unpublished
 COMMENT Contact: Genoscope

Genoscope - Centre National de Sequencage
BP 191 91006 Evry cedex - France
Email: seque@genoscope.cns.fr, Web: www.genoscope.cns.fr
Library was constructed by Life Technologies, a division of
Invitrogen. This sequence belongs to sequence cluster 2753.r For
more information about this cluster, see
http://www.genoscope.cns.fr/
cgi-bin/cluster.cgi?seq=CS0DM05AD03QPLcluster=2753.r. Contact :
Peng Liang Email: fliang@lifetech.com URL: <http://fulllength.invitrogen.com/> Invitrogen Corporation 1600
Faraday Avenue Genoscope sequence ID : CS0DM05AD03QPL.

FEATURES

source

1. .983
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0DM051G05"
/issue_type="FETAL LIVER"
/dev_stage="fetal"
/clone_lib="Homo sapiens FETAL LIVER"
/note="Organ: liver; Vector: pCMVSPORT 6; 1st strand cDNA
was primed with a NotI-oligo(dt) primer. Five prime end
enriched, double-strand cDNA was digested with Not I and
cloned into the Not I and EcoRV sites of the pCMVSPORT 6
vector. Library was not normalized."

BASE COUNT 286 a 165 c 209 g 313 t 10 others

ORIGIN

Query Match 36.8%; Score 606.8; DB 13; Length 983;
Best Local Similarity 80.7%; Pred. No. 2,1e-117;
Matches 736; Conservative 9; Mismatches 166; Indels 5; Gaps 3;

15 AAGCATTCAGTTCAGATGATGCTATGAAATGAGCTTCCTCTCTGATATACG 74
65 AAGATTCAGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 124
75 TGAGCTTACTTACTTACTTACTTACTTACTTACTTACTTACTTACTTACTTACT 134
125 TCAATTTTACTTACTTACTTACTTACTTACTTACTTACTTACTTACTTACTTACT 184
135 GCCATCGATGATATATGAAGCAATCTCGATGAACTTTGTCAGAGAGTCTAGAG 194
185 GCCATTCGATTAATATGATGATGATGATGATGATGATGATGATGATGATGATG 244
195 CTGATTCGATCTTCTGATCTTCTGATCTTCTGATCTTCTGATCTTCTGATCTT 254
245 CTGATTCGATCTTCTGATCTTCTGATCTTCTGATCTTCTGATCTTCTGATCTT 304
255 TTGAAGTTATCTGATCTTCTGATCTTCTGATCTTCTGATCTTCTGATCTTCTG 314
305 TAGAAGTTATCTGATCTTCTGATCTTCTGATCTTCTGATCTTCTGATCTTCTG 364
315 TTAAGATGAGCA--GAACCTTCAAAAGACATTTTGGTCAATTTTTCACAGATAC 371
365 TCATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 424
372 AAGAAATCATGTGACATTTTATGACATCTTAAAGATTCTTAAAGATATAGTTCAA 431
425 AAGAAATGTGTGGAAATATATATGACATTAACAAGCTCTTAAAGATGACATTTTGA 484
432 ATAAGAAATTAATAAGAACTACAGAGAGTCAAGATTTGATGTTCTTGGACAGTCTG 491
485 ATAAGAAATTAATAAGAACTACAGAGAGTCAAGATTTGATGTTCTTGGACAGTCTG 544
492 TTTTCCCTTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 551
545 TTAATCCCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 604
552 GCTTCTCTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 611
605 GATTCCTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 664
612 ATGTCCTGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 671

665 ATGACCTGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 724
672 ATATGATCTATGCTCTTATTTGATTTTGGTTCCAAATTTTGAATGATGAAGAGGCG 731
725 ATATGATCTATGCTCTTATTTGATTTTGGTTCCAAATTTTGAATGATGAAGAGGCG 784
732 ATGAGTCTCTGAGGAAGTCTGAGGAAGCCACTAGCTATGATGATGATGATGATGATG 791
785 ACCAGTTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 844
792 CTGACATATGCTTATTTGAACT-ACCTGATTTTCAATTTTCTCAACCACTTTTCA 850
845 CTGAATATGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 904
851 AATGTTGATGCTGTTGAGGATCTGCTCACTCACTCACTCACTCACTCACTCACTCA 910
905 AATGTTGATGCTGTTGAGGATCTGCTCACTCACTCACTCACTCACTCACTCACTCA 963
911 GAAGATTTGCTCAGA 926
964 GAAGATTTGCTCAGA 979

RESULT 10
LOCUS B0713091 960 bp mRNA linear EST 16-JUL-2002
DEFINITION AGENCOURT_8295152 NIH_MGC_129 Mus musculus cDNA clone IMAGE:6309767
ACCESSION B0713091
VERSION B0713091.1 GI:21851990
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 960)
NIH-MGC <http://mgc.ncl.nih.gov/>
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished
Contact: Robert Straube, Ph.D.
Email: c9abps-remail.nih.gov
Tissue Procurement: Susan L. Sullivan, PhD.
CDNA Library Preparation: Resgen, Invitrogen Corp.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
<http://image.lnl.gov>
Plate: LLM13730 row: p column: 24
High quality sequence stop: 737.
Location/Qualifiers
1. .960
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="IMAGE:6309767"
/lab_host="DH10B (phage-resistant)"
/note="Organ: olfactory epithelium; Vector:
pCMV-SPORT6.1.cdb; Site 1: EcoRV; Site 2: NotI; Cloned
unidirectionally. Primer: Oligo dt. Average insert size
2.2 Kb. Constructed by Resgen, Invitrogen Corp. Note: this
is a NIH MGC Library."

FEATURES

source

1. .960
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="IMAGE:6309767"
/lab_host="DH10B (phage-resistant)"
/note="Organ: olfactory epithelium; Vector:
pCMV-SPORT6.1.cdb; Site 1: EcoRV; Site 2: NotI; Cloned
unidirectionally. Primer: Oligo dt. Average insert size
2.2 Kb. Constructed by Resgen, Invitrogen Corp. Note: this
is a NIH MGC Library."

BASE COUNT

273 a 204 c 237 g 246 t

ORIGIN

Query Match 36.4%; Score 600.6; DB 13; Length 960;
Best Local Similarity 80.7%; Pred. No. 4.2e-116;
Matches 713; Conservative 0; Mismatches 169; Indels 1; Gaps 1;

745 TGAAGTTTGAAGAACCACTGATCTGATGATGATGATGATGATGATGATGATGATG 804

Db 36 TAAAGTACAGAGAACCCACCACTGTGTAGACTATGGGAAAGCTGAGATTGGCT 95
 Qy 805 TATTGGAACCTACGCGGATTTTCAATTTCCACCCACTCTTACCAATGTTGATCGT 864
 Db 96 AATGGAACCTATTGGGATTTTGAATTTCTGCTATATTATACCAATTTTGGATTGT 155
 Qy 865 TGGAGACTCCACTGCAACCTGCGCAACCCCTACCGAAGAAATGGAAGATTGTCCA 924
 Db 156 GGGAGACTGCACTGCAACCTGCGCAACCCCTTACTAGAAATGGAAGATTGTCCA 215
 Qy 925 GAGCTTGGAGAAATGTTGTTGTTTCTCTGCGGTGATGTCAGTAAACAGCTC 984
 Db 216 GACCTCAGGAGGAACATGTTATTTGTTGTTTCTCTGCGGTCAATATGTCAAATACCTGAC 275
 Qy 985 AGAAGAAAGGCGCAATGTAATGTCAGCCCTTGGCCCAAGATCCCAAGAAAGTTCTGTG 1044
 Db 276 AGATGAAAGGCGCAATGTCATTTGCTCAGCCCTTGGCCCAAGATCCCAAGAAAGTTCTGTG 335
 Qy 1045 GAGATTTGATGGAAATAACAGATACTTTAGACTCAATCTCGGCTGTACAGTGGAT 1104
 Db 336 GCGATCAAGAAAGAAATACCAAGACATTTAGATCCAAATCAAGACTGTTGATGGAT 395
 Qy 1105 ACCCGAATGATCTCTTGTGTCATCCCAAAACCAAGCTTTTATCACTCATGTGGAT 1164
 Db 396 TCCTCAGATGATCTTCTTGGACATCCCAAAACCAAGCTTTTATCACTCATGTGGAT 455
 Qy 1165 GAAATGGATCTTATGAACTTTTACATGGGTCTCTTATGTTGGAGATTTCCCATTTGG 1224
 Db 456 AATATGAAATCTATGAGGCTTATTTACATGGATTTCTTATGTTGGAGATTTCCCATTTGG 515
 Qy 1225 TGATGAGCTGTGATACATAGCTCACATGAGGCGCAAGAGAGAGCTGTAGAAATTAATCTT 1284
 Db 516 TGACCAAGCTGACCAATTTGCTCACAATGAGGCGCAAGAGAGAGAGCTGTAGAAATTAATCTT 575
 Qy 1285 CAAAATCTATGACAAAGATTTTACTGAGGCTTTGAGAAACATGATTAATCCGATCTCTC 1344
 Db 576 GAAACGATGACAAAGCTCAGATCTCTCAATGCTCTGAGAAACATGATTAATGACCATTC 635
 Qy 1345 TTTATTAAGAAATGCTATGAAATTTTCAAGATTTTCAAGATTTTCAAGATTTTCAAGATTTT 1404
 Db 636 TTTATTAAGAAATGCTATGAAATTTTCAAGATTTTCAAGATTTTCAAGATTTTCAAGATTTT 695
 Qy 1405 AGATGAGAGAGCTCTCTGAGATGAGTTTGTCTGAGGCGCAAGAGAGAGAGAGAGAGAGAGAG 1464
 Db 696 GAGAGAGAGAGCTCTCTGAGATGAGTTTGTCTGAGGCGCAAGAGAGAGAGAGAGAGAGAGAG 755
 Qy 1465 ATCAGCTGCGCCATGACCTGAGCTGTTCCAGACTACTCTATGATGATGATGATGATGATGAT 1524
 Db 756 TGTGCGAGGCGCATGACCTGAGCTGTTCCAGTACACTCTCTGATGATGATGATGATGATGAT 815
 Qy 1525 GGTGACCTGTGTGCGCACTGCTATATCTGTTTCAAAATGTTTATTTT-TCCTGTC 1583
 Db 816 GCTGGGCTGTGTGCGCACTGCTATATATTTGTTTCAAAATGCTGTTTGTATTAATTTTC 875
 Qy 1584 AAAAAATTAATAAATCTAGAAATGAAAGAGGAATATGAT 1626
 Db 876 AAAAAATTTGCTTAAGACAGAAAGAGAAATTAATGATTTGATTT 918

RESULT 11
 B0942104 1004 bp mRNA linear EST 21-AUG-2002
 LOCUS AGENCOURT 8763113 NIH_MGC_129 Mus musculus cDNA clone IMAGE:6313348
 DEFINITION 5', mRNA sequence.
 ACCESSION B0942104
 VERSION B0942104.1 GI:22357582
 KEYWORDS EST.
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 REFERENCE 1 (bases 1 to 1004)
 NIH-MGC http://mgs.nci.nih.gov/.

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: gsgabs-remail.nih.gov
 Tissue Procurement: Susan L. Sullivan, PhD.
 CDNA Library Preparation: Resgen, Invitrogen Corp.
 DNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
 DNA Sequencing by: Agencourt Bioscience Corporation
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LNL at:
 http://image.lnl.gov
 Plate: L1M13740 row: F column: 05
 High quality sequence stop: 656.
 Location/Qualifiers
 1..1004
 /organism="Mus musculus"
 /mol_type="mRNA"
 /db_xref="taxon:10090"
 /clone="IMAGE:6313348"
 /lab_host="DH10B (phage-resistant)"
 /clone_lib="NIH MGC 129"
 /note="Organ: olfactory epithelium; Vector: pCMV-Sport6.1.cdb; Site: 1. EcorV; Site 2: NotI; Cloned unidirectionally. Primer: Oligo dt. Average insert size 2.2 kb. Constructed by Resgen, Invitrogen Corp. Note: this is a NIH MGC Library."
 BASE COUNT 286 a 219 c 249 g 249 t 1 others
 ORIGIN
 Query Match 36.4%; Score 599.8; DB 13; Length 1004;
 Best Local Similarity 80.0%; Pred. No. 6.2e-116;
 Matches 718; Conservative 0; Mismatches 178; Indels 2; Gaps 1;
 Qy 721 GAAGAATGGATAGCTTCAAGTGAAGTTCTGAAGAGCCCACTGATATGAGAC 780
 Db 46 GAACAGTGGATTCCTATTATTAAGCTTTGAGAGAGCCCACTGATGAGAC 105
 Qy 781 AATGGCAAAAGCTGACATATGCTTATTCGAACTACTGGATTTTCAATTTCTCAACC 840
 Db 106 TATGGGAAGAGCTGAGATTTGGCTAATGCAACTAATTTGGATTTTGAATTTCTCGTCC 165
 Qy 841 ACTCTTACCAAAATGTTGATGTTGTTGAGAGACTCACTGCAAACTGCAAACTGCAAACTGCAAA 900
 Db 166 ATATTTACCGAATTTTGAATTTGTTGAGAGAGCTCACTGCAAACTGCAAACTGCAAACTGCAAA 225
 Qy 901 GAAGAAATGGAAGATTTTGTCCAGAGCTCTGAGAAATGTTGTTGTTGTTGTTGTTGTTGTTGTT 960
 Db 226 TAAAGAAATGGAAGATTTTGTCCAGAGCTCTGAGAAATGTTGTTGTTGTTGTTGTTGTTGTTGTT 285
 Qy 961 GGGTGCATGCTGCTAAACAGTCAAGAAAGAGGCGCAATGTAATGATCAGGCTTGC 1020
 Db 286 GGGGTCAATGCTCAAAAACCTGACAGATGAAAGGCGCAATGTAATGATCAGGCTTGC 345
 Qy 1021 CAAGATCCCAAAAAGTTCTGTGAGATTTGATGGAAATTAACCAATTTTACATGAGACT 1080
 Db 346 CCAAGATCCCAAAAAGTTCTGTGAGATTTGATGGAAATTAACCAATTTTACATGAGACT 405
 Qy 1081 CAATATCTGCTGTCAAGTGTATACCCCAAGATTTTGTGATATCCCAAAACCA 1140
 Db 406 CAATATCAAGATCTGTTGATGTTGATTTCTCAAGATTTTGTGATATCCCAAAACCA 465
 Qy 1141 AGCTTTTATCACTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1200
 Db 466 AGCTTTTATCACTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 525
 Qy 1201 TATGTTGAGATTTCCCATATTTTGTGATCAGCTTGTATTAATAGCTCATGAGAGCCAA 1260
 Db 526 TATGTTGAGATTTCCCATATTTTGTGATCAGCTTGTATTAATAGCTCATGAGAGCCAA 585
 Qy 1261 AGGAGCAGCTGATGAAATTAATCTTCAAAATCTATGACAGCAATTTACGAGGCTTT 1320
 Db 586 AGGAGCAGCAGTGTGAGGATGAAATGAAATGAAATGAAATGAAATGAAATGAAATGAAATGAAATGAA 645

QY 1321 GAGAACACTCATTACCGATTCTCTTATTAAGAGATGCTATGAGATTATCAAGATTCA 1380
DB 646 GAGAAACAGTTCATCATGAGCCATCTTATTAAGAGATGCTATGAGATTATCAAGATTCA 705
QY 1381 CCATGATTAACCTGTAAGAGCCCTTATGATGAGCAGTCTTCTGATCGAGTTTGTATCGG 1440
DB 706 CCATGACCAAGCAATGAAAGCCCTTATGATGAGCAGTCTTCTGATCGAGTTTGTATCGG 765
QY 1441 CCACAAAGGAGCAGCAGCAGTCTGATGAGTCCCATGAGTCTTCTGATCGAGTTTGTATCGG 1500
DB 766 TCACAAAGGAGCAGCAGCAGTCTGATGAGTCCCATGAGTCTTCTGATCGAGTTTGTATCGG 825
QY 1501 CTCTATGATGATGATGAGTCTGATGAGTCTGATGAGTCTTCTGATCGAGTTTGTATCGG 1558
DB 826 CTCTCTGATGAGTATGAGTCTGATGAGTCTGATGAGTCTTCTGATCGAGTTTGTATCGG 885
QY 1559 ACAAATGTTTATTTATTTCTGTCAGAAATTTAATAAAGTAAAGATGAAAGA 1616
DB 886 CAAAGAGGTGTTGTTTATTTCAAAAGTGTGTAAGACCGGAAAGAAAAA 943

RESULT 12
BG562901 595 bp mRNA linear EST 10-APR-2001
LOCUS 602581752F1 NIH_MGC_76 Homo sapiens cDNA clone IMAGE:4709436 5',
mRNA sequence.
ACCESSION BG562901
VERSION BG562901.1 GI:13570553
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 595)
AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.
JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT Unpublished
Contact: Robert Strausberg, Ph.D.
Email: cgabbs-remail.nih.gov
Tissue Procurement: CLONETECH Laboratories, Inc.
CDNA Library Preparation: CLONETECH Laboratories, Inc.
DNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNLN at:
<http://image.llnl.gov>
Plate: LNCMI546 row: h column: 13
High quality sequence stop: 595.
Location/Qualifiers
1..595
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:4709436"
/lab_host="DH10B (T1 phage-resistant)"
/clone_lib="NIH MGC 76"
/note="Organ: liver; Vector: pDNR-LIB (Clontech); Site 1:
SfiI (ggcgccgcggcc); Site 2: SfiI (ggcgccatggcc); 5' and
3' adaptors were used in cloning as follows: 5' adaptor
sequence: 5'-CACGCCATTAATGACC-3' and 3' adaptor sequence:
5'-ATTCTAGAGCCGAGGCGGCGACATG-dt(30)BN-3' (where B = A,
C, or G and N = A, C, G, or T). Average insert size 1.85
kb (range 1.0-4.0 kb). 15/15 colonies contained inserts
by PCR. This library was enriched for full-length clones
and was constructed by Clontech Laboratories (Palo Alto,
CA). Note: this is a NIH MGC Library."

BASE COUNT 175 a 115 c 128 g 177 t
ORIGIN

Query Match 35.2%; Score 580.8; DB 10; Length 595;
Best Local Similarity 99.7%; Pred. No. 6.1e-112;
Matches 582; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AGCAACTGGAACAAACGATTGATTCATCAGAGATGCTATGAAATGAGCTTCAGCTCT 60
DB 12 AGCAACTGGAACAAACGATTGATTCATCAGAGATGCTATGAAATGAGCTTCAGCTCT 71
QY 61 TCTCTGATACAGCTGAGCTGTTATCTTACCTCTGGAGTTTGGAAAAGTCTGTGTG 120
DB 72 TCTGCTGATACAGCTGAGCTGTTATCTTACCTCTGGAGTTTGGAAAAGTCTGTGTG 131
QY 121 GCCCAGGAATTCAGCCATGAGTGAATATTAAGACAAATCTGGATGAACCTGTCCAGAG 180
DB 132 GCCCAGGAATTCAGCCATGAGTGAATATTAAGACAAATCTGGATGAACCTGTCCAGAG 191
QY 181 AGGTCATGAGTGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 240
DB 192 AGGTCATGAGTGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 251
QY 241 ATCTACTCTTAAATTTGAAGTTTATCTGATCTTTTAACTTAACTGAGTTTGAAGATAT 300
DB 252 ATCTACTCTTAAATTTGAAGTTTATCTGATCTTTTAACTTAACTGAGTTTGAAGATAT 311
QY 301 TATCAAGCAGCTGGTTTGAAGATGAGGAGAACTTCCAAAGACATTTGTGTCATATTT 360
DB 312 TATCAAGCAGCTGGTTTGAAGATGAGGAGAACTTCCAAAGACATTTGTGTCATATTT 371
QY 361 TTCAAGATGACAAAGAAATCATGTGACATTTAATGACATTTAAGAAAGTTCTGTAAAGA 420
DB 372 TTCAAGATGACAAAGAAATCATGTGACATTTAATGACATTTAAGAAAGTTCTGTAAAGA 431
QY 421 TATAGTTTCAATTAAGAACTTATGAAGAACTACAGAGTCAAGATTTGATGTTCT 480
DB 432 TATAGTTTCAATTAAGAACTTATGAAGAACTACAGAGTCAAGATTTGATGTTCT 491
QY 481 TGCAGATGCTGTTTCCCTTTTGTGATGCTGCTGCGGAGTACTTAAATACCTTTGT 540
DB 492 TGCAGATGCTGTTTCCCTTTTGTGATGCTGCTGCGGAGTACTTAAATACCTTTGT 551
QY 541 CTACAGCCTCGCTTCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 584
DB 552 CTACAGCCTCGCTTCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 595

RESULT 13
BG428781 814 bp mRNA linear EST 14-MAR-2001
LOCUS 602500860F1 NIH_MGC_75 Homo sapiens cDNA clone IMAGE:4614566 5',
mRNA sequence.
ACCESSION BG428781
VERSION BG428781.1 GI:13335287
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 814)
AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.
JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT Unpublished
Contact: Robert Strausberg, Ph.D.
Email: cgabbs-remail.nih.gov
Tissue Procurement: CLONETECH Laboratories, Inc.
CDNA Library Preparation: CLONETECH Laboratories, Inc.
DNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNLN at:
<http://image.llnl.gov>
Plate: LNCMI366 row: g column: 15
High quality sequence stop: 761.
Location/Qualifiers
1..814
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"

FEATURES
SOURCE

```

/clone="IMAGE:4614566"
/lab host="DH10B (T1 phage-resistant)"
/clone.lib="NIH_MGC_75"
/notes="Organ: Kidney; Vector: pDNr-LIB (Clontech); Site_1:
5'fl (ggcgccctcgcc); Site_2: 5'fl (ggcattagcc); 5' and
3' adaptors were used in cloning as follows: 5' adaptor
sequence: 5'-CACGCCATTATGGCC-3' and 3' adaptor sequence:
5'-ATTCTAGAGCGCGCGCGCGCATGCG-dt(30)BN-3' (where B = A,
C, or G and N = A, C, G, or T). Average insert size 1.65
kb (range 0.5-4.0 kb). 15/15 colonies contained inserts
by PCR. This library was enriched for full-length clones
and was constructed by Clontech Laboratories (Palo Alto,
CA). Note: this is a NIH MGC Library."

BASE COUNT      224 a      167 c      182 g      241 t
ORIGIN

Query Match      34.9%; Score 576.4; DB 10; Length 814;
Best Local Similarity 85.8%; Pred. No. 5.3e-111;
Matches 688; Conservative 0; Mismatches 106; Indels 8; Gaps 4;

QY 453 TACAGAGTCAAGATTGATGTTCTTTCAGAGTCTTTCCCTTGGTGGAGCTGC 512
Db 1 TACAAGAGTCAAGATTGATGTTCTTTCAGAGTCTTTCCCTTGGTGGAGCTGC 60
QY 513 TGGCCGAGTACTTAAATACCTTTGTCTACAGCCTCCGCTTCTCTCGGCTACGCA 572
Db 61 TGGCTGAGCTATTAACTACCTTTGTGTACAGTCTACGCTTCTCTCGGCTACCTT 120
QY 573 TTGAAGAGTATGAGAGAGCTTCTGTTCCCTCTCTTCTGATGCTGTTATGTCAG 632
Db 121 TTGAAGAGTATGAGAGAGTATTTTCCCTCTCTCTCTCTGATGTTATGTCAG 180
QY 633 AACTAGAGTCAAGTCTTCTTATGAGAGGCTTAAATATGATGATGCTTATTT 692
Db 181 AATTAAGTATCAAGTCTTCTTATGAGAGGCTTAAATATGATGATGCTTATTT 240
QY 693 TTGAATTTTGGTTCCTTCAATATTTGACATGAGAGTGGAGTCACTTCAAGTGAATTC 752
Db 241 TTGACTTTTGGTTCCTTCAATATTTGACATGAGAGTGGAGTCACTTCAAGTGAATTC 300
QY 753 TAGAAGAGCCACTACGTTATCTGAGACATGAGCAAAAGCTGACATATGCTTATTCGA 812
Db 301 TAGAAGAGCCACTACGTTATCTGAGACATGAGCAAAAGCTGACATATGCTTATTCGA 360
QY 813 ACTATCTGGATTTTCAATTTCTCTCCACCTCTTACCAATGTTAGTTCGTTGAGAGAC 872
Db 361 ACTCTGGAATTTTCAATTTCTCTCCACCTCTTACCAATGTTAGTTCGTTGAGAGAC 419
QY 873 TCCACTGCAAACTCTGCAAAACCTCTACCGAAGAAATGGAAGTTTGTCCAGACTCTG 932
Db 420 TCCACTGCAAACTCTGCAAAACCTCTGCGGCTCTGAGAAATGGAAGTTTGTGACAGACTCTG 479
QY 933 GAG-AAAATGTTGTTGTTGTTTCTCTGCGGCTGAGTGTCACTACAGCTCAAGAGA 991
Db 480 GAGCAAAAATGTTGTTGTTGTTTCTCTGCGGCTCAATGTCACTACAGCTCAAGAGA 539
QY 992 AGGCGCAATGTAATGCACTACAGCCTTGGCAAGATCCCAAAAGTTCTGTGAGATT 1051
Db 540 AGGCGCAATGTAATGCACTACAGCCTTGGCAAGATCCCAAAAGTTCTGTGAGATT 598
QY 1052 GATGGGAATTAACAGATCTTATAGACTCAATACCTGCTGATACAGTGAATCCCGAG 1111
Db 599 GATGGGAATTAACAGATCTTATAGACTCAATACCTGCTGATACAGTGAATCCCGAG 658
QY 1112 AATGATCTTCTTGTGCTATCCCAAAACCAAGCTTTTATCACTCATGTGGAATG-----A 1166
Db 659 AATGATCTTCTTGTGCTATCCCAAAACCAAGAGCTTTTATTAATCATGTGTTGAGAGCA 718
QY 1167 ATGGATCTATGAGATATTTACATGGGGTCCCTATGTTGGAAGTTCCCATATTTGGTG 1226
Db 719 TGGGATCTATGAGAGATCTATCATGAGGAATCTATGTTGGAATTCATTTGTTGGG 778
QY 1227 ATCAGTTGATTAACATAGCTCA 1248

```

```

Db 779 ATCAACCTGTAAATGCTCA 800

|||||
RESULT 14
BQ925596 927 bp mRNA linear EST 20-AUG-2002
LOCUS
DEFINITION
AGENCOURT 8688479 NIH_MGC_129 Mus musculus cDNA clone IMAGE:6311424
5', mRNA sequence.
BQ925596
ACCESSION
BQ925596.1 GI:22340627
VERSION
BQ925596.1
KEYWORDS
SOURCE
Mus musculus (house mouse)
ORGANISM
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
1 (bases 1 to 927)
AUTHORS
NIH-MGC http://mgi.nci.nih.gov/.
TITLE
National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL
Unpublished
COMMENT
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: Susan L. Sullivan, Ph.D.
CDNA Library Preparation: ResGen, Invitrogen Corp
DNA sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
http://image.llnl.gov
Plate: L14M13735 Row: E Column: 01
High quality sequence stop: 635.
Location/Qualifiers
1..927
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="IMAGE:6311424"
/lab host="DH10B (phage-resistant)"
/clone.lib="NIH_MGC_129"
/notes="Organ: olfactory epithelium; Vector:
PCMV-SPORE6.1.ccd; Site_1: EcoRV; Site_2: NotI; Cloned
unidirectionally. Primer: Oligo dt. Average insert size
2.2 kb. Constructed by ResGen, Invitrogen Corp. Note: this
is a NIH MGC Library."

BASE COUNT      262 a      201 c      230 g      231 t      3 others
ORIGIN

Query Match      34.9%; Score 575.4; DB 13; Length 927;
Best Local Similarity 78.3%; Pred. No. 8.6e-111;
Matches 702; Conservative 0; Mismatches 193; Indels 2; Gaps 1;

QY 745 TGAAGTTCTAGAGAGCCCACTAGCTTATCGAACAATGGCAAAAGCTGACATATGGCT 804
Db 28 TAAAGTACAGAGAGCCCACTAGCTGTGGAACATGAGGAAAGTGAATTTGGCT 87
QY 805 TATTCGAATACACTGAGATTTTCAATTTCTCACTCCCACTCTTACCAATGTTGAGTTCT 864
Db 88 AATGGAACCTATGAGATTTTGAATTTCTCTGCTGCTATTTACGAAATTTGAGTTGT 147
QY 865 TGAAGACTGCACTGCAAACTGCGCAAAACCTTACCGAAGAAATGGAAGTTTGTCCA 924
Db 148 GGAAGACTGCACTGCAAACTGCGCAAAACCTTACCTTAAGAAATGGAAGTTTGTCCA 207
QY 925 GAGCTCTGAGAAAATGTTGTTGTTGTTTCTCTGCGGCTGATGCTACATACAGTCTC 984
Db 208 GACCTCAGGGGAAATGTTGTTGTTGTTTCTCTGCGGCTGATGCTACATACAGTCTC 267
QY 985 AGAAGAAAGGCAATGATTTGATGATGAGCCCTTGCAAGATCCCAAAAGTTCTGTG 1044
Db 268 AGATGAAAGGCAATGATTTGATGATGAGCCCTTGCAAGATTTCTCAAGAGTTTGTG 327
QY 1045 GAGATTTGATGAGAAATAACAGATCTTATAGACTCAATACCTGCTGTAAGTGGAT 1104

```

```

Db      328 GCGATACAAAGAAAAGATACAGACATTAAGATCCAAATGCAAGACTGTTTGATGGAT 387
Qy      1105 ACCCCAGATGATCTTCTTGGTCAATCCAAAACCAAGCTTTTCACTCATGATGGAT 1164
Db      388 TCTTCAGATGATCTTCTTGGTCAATCCAAAACCAAGCTTTTATACCCATGGTGGAC 447
Qy      1165 GAATGGATCTATGAAGCTATTTACCATGGGGTCCCTATGTTGGAGTTCCTATTTGG 1224
Db      448 AATATGATCTATGAGCTATTTACCATGGATCCCTATGTTGGAGTTCCTATTTGG 507
Qy      1225 TGATCAGCTTATACATAGCTACATGAGCCCAAGAGACAGCTGTGAATTAATTT 1284
Db      508 TGACACAGCTTACCAATTTGCTCATGTAGAGCCCAAGAGACAGCTGTGAATTAATTT 567
Qy      1285 CAAACTATGACCAAGCAAGATTTACTGAGGGCTTTGGAACAGTCAATTCAGTTCCTC 1344
Db      568 GAACACATGACCAAGCTTCAATCTGCTCAATGCTTGAAGACAGTCAATGAGCCATTC 627
Qy      1345 TTATTAAGAGATGCTATGATTAATCAAGATTCACCATGATCAACCTGTAAAGCCCT 1404
Db      628 TTATTAAGAGATGCTATGATTAATCAAGATTCACCATGATCAACCTGTAAAGCCCT 687
Qy      1405 AGATGACAGCTCTTCTGATGAGTTGTGATGCGCCCAAGAGACCAAGCTGCG 1464
Db      688 GGACCGAGCAGCTCTTCTGATGAGTTGTGATGCGCCCAAGAGACCAAGCTGCG 747
Qy      1465 ATCAGCTGCGCAGCTCACTGCTGCTGACAGCTCACTGATGATGATGATGATGATGAT 1524
Db      748 TGTGCGACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 807
Qy      1525 GCTGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1582
Db      808 GCTGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 867
Qy      1583 CAAATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1639
Db      868 TTCCAAAAGTTGGTTAGAACCGGGGAAAGAAAACCGTGGAGCTTAA 924

RESULT 15
ID      HSM087819 standard; RNA; EST; 588 BP.
XX      AC      BX494989;
XX      SV      BX494989.1
XX      DT      09-MAY-2003 (Rel. 75, Created)
XX      DT      09-MAY-2003 (Rel. 75, Last updated, Version 1)
DE      Homo sapiens mRNA; EST DKFZp779L0816_r1 (from clone DKFZp779L0816)
XX      EST; expressed sequence tag.
XX      KM      Homo sapiens (human)
XX      OS      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia;
XX      OC      Euteleostomi; Primates; Catarrhini; Hominiidae; Homo.
XX      RN      [1]
XX      RP      1-588
RA      Wambutt R., Heubner D., Mewes H.W., Weil B., Amid C., Osanger A., Fobo G.,
RA      Han M., Wiemann S.;
RT      Submitted (07-MAY-2003) to the EMBL/GenBank/DBJ databases.
RL      MITS, Ingolstaedter Landstr.1, D-85764 Neunherberg, GERMANY
CC      This is the 5' sequence of the clone insert
CC      Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
CC      Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;
CC      Sequenced by AGOWA (Berlin/Germany) within the cDNA sequencing
CC      Consortium of the German Genome Project.
CC      No 5' sequence available.
CC      This clone (DKFZp779L0816) is available at the RZPD in Berlin.

```

```

CC      Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6,
CC      14059 Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de
XX      Key      Location/Qualifiers
FH      source      1. 588
FT      /db_xref="taxon:9606"
FT      /mol_type="mRNA"
FT      /organism="Homo sapiens"
FT      /clone_lib="DKFZp779L0816"
FT      /clone_1lib="779 (synonym: hnccl). Vector pSport1_Sfi; host
FT      DH10B; sites SfiIA + SfiIB"
FT      /dev_stage="fetal"
FT      /tissue_type="liver"
XX      SQ      Sequence 588 BP; 170 A; 114 C; 127 G; 177 T; 0 other:

Query Match      34.8%; Score 573.4; DB 2; Length 588;
Best Local Similarity 99.8%; Pred. No. 2.2e-110;
Matches 574; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy      1 AGCACTGGAAGAAACACATTTGATTCATGAGATGCTATGAAATGGAATGACTTCACTCT 60
Db      14 AGCAACTGGAAGAAACACATTTGATTCATGAGATGCTATGAAATGGAATGACTTCACTCT 73
Qy      61 TCTCTGATACAGCTGAGCTGTTACTTACTTACTGCTGGAGTGTGGAAGGCTGCTGCTG 120
Db      74 TCTCTGATACAGCTGAGCTGTTACTTACTTACTGCTGGAGTGTGGAAGGCTGCTGCTG 133
Qy      121 GCCCAGAAATTCAGCCAGCTGATGATTAAGACAAATCCTGATGAATCTGTCAGAG 180
Db      134 GCCCAGAAATTCAGCCAGCTGATGATTAAGACAAATCCTGATGAATCTGTCAGAG 193
Qy      181 AGCTATGAGGTGATGCTATTTGGCATCTTCAGCTTCATTTCTTTGATCCCAACAGCCC 240
Db      194 AGCTATGAGGTGATGCTATTTGGCATCTTCAGCTTCATTTCTTTGATCCCAACAGCCC 253
Qy      241 ATCTACTCTTAAATTTGAAGTTTATCTGATCTTAACTTAACTGAATTTGAGATTTAT 300
Db      254 ATCTACTCTTAAATTTGAAGTTTATCTGATCTTAACTTAACTGAATTTGAGATTTAT 313
Qy      301 TATCAAGCAGCTGTTAAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 360
Db      314 TATCAAGCAGCTGTTAAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 373
Qy      361 TTCAAGTACAGAAATTCATGAGATTTAATGAATTAATGAATTAATGAATTTCTTAAGGA 420
Db      374 TTCAAGTACAGAAATTCATGAGATTTAATGAATTAATGAATTTAATGAATTTCTTAAGGA 433
Qy      421 TATAGTTTCAAAATTAAGAACTTATGAAGAACTACAGAGTCAAGATTTGATGTTGTTCT 480
Db      434 TATAGTTTCAAAATTAAGAACTTATGAAGAACTACAGAGTCAAGATTTGATGTTGTTCT 493
Qy      481 TGCAATGCTGTTTCCCTTTGGTGAAGTGTGAGCTGCGGAGTTACTTAAATACCTTTGT 540
Db      494 TGCAATGCTGTTTCCCTTTGGTGAAGTGTGAGCTGCGGAGTTACTTAAATACCTTTGT 553
Qy      541 CTACAGCTTCGCTTCTCTCTGCTGCTAGCAATTG 575
Db      554 CTACAGCTTCGCTTCTCTCTGCTGCTAGCAATTG 588

```

Search completed: December 7, 2003, 07:38:01
 Job time : 2473 secs

THIS PAGE BLANK (USPTO)

THIS PAGE BLANK (USPTO)